U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

Requestor's Name: RAILEY

Serial Number: 08/000,716

Date: 21 APRIL 1993 Phone: 308-0281 Art Unit: 1804

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

PLEASE SEARCH CLAIM 11, IT IS DIRECTED TO HIV SEQUENCES. VIRUS KNOWN AS LYMPHADENOPATHY ASSOCIATED VIRUS (LAV) OR HUMAN T CELL LYMPHOTROPIC VIRUS TYPE III (HTLV-III). NOW CALLED HUMAN IMMUNODEFICIENCY VIRUS TYPE I. I NEED ONLY THE "HITS" CLOSEST T: [redacted] HOMOLOGY.

U.S. PAT. & TM OFF.
93 APR 21 PM 5:48
BIOTECH/CHEMICAL
RECEIVED

4-181

STAFF USE ONLY

Date completed: 04-26-93
Searcher: Beverly@4994
Terminal time: 47
Elapsed time: 18
CPU time:
Total time: 57
Number of Searches:
Number of Databases: 2

Search Site
- STIC
- CM-1
- Pre-S

Type of Search
- N.A. Sequence
- A.A. Sequence
- Structure
- Bibliographic

Vendors
- IG Suite
- STN
- Dialog
- APS
- Geninfo
- USDC
- DARC/Questel
- Other

Railey
08/060716

=> fil reg; s ggggg[redacted]ggaagggctaattcactcccaa/sq[redacted]
FILE 'REGISTRY' ENTERED AT 14:41:40 ON 26 APR 9[redacted]
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 1993 American Chemical Society (ACS)

STRUCTURE FILE UPDATES: 23 APR 93 HIGHEST RN 147199-92-6
DICTIONARY FILE UPDATES: 25 APR 93 HIGHEST RN 147199-92-6

EXCLUDE SEARCH OF COMPLEMENTARY STRAND Y/(N)?:.
L1 35 GGGGGACTGGAAGGGCTAATTCACTCCCAA/SQSN Seq. claim 11

=> fil ca; s l1
FILE 'CA' ENTERED AT 14:42:04 ON 26 APR 93
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 1993 AMERICAN CHEMICAL SOCIETY (ACS)

FILE COVERS 1967 - 13 Apr 93 (930413/ED) VOL 118 ISS 16.
For OFFLINE Prints or Displays, use the ABS or ALL formats to obtain abstract graphic structures. The AB format DOES NOT display structure diagrams.

L2 9 L1

=> d 1-9 .beverly; sel hit l2 1-9 rn

L2 ANSWER 1 OF 9 COPYRIGHT 1993 ACS
AN CA117(21):206366s
TI Molecular clones of HIV-1 strains MN-ST1 and BA-L and preparation of vaccines with antigenic proteins of these strains
SO PCT Int. Appl., 55 pp.
AU Reitz, Marvin S., Jr.; Franchini, Genoveffa; Markham, Phillip D.; Gallo, Robert C.; Lori, Franco C.; Popovic, Mikulas; Garnter, Suzanne
AI WO 91-US7611 17 Oct 1991
PI WO 9206990 A1 30 Apr 1992
PY 1992
AB HIV-1 strain MN-ST1 cDNA and a HindIII fragment of strain BA-L cDNA are cloned and sequenced. Plasmids for expression of infectious viruses or env protein were prepd. Restriction maps of MN-ST1 prophage cDNA and of the cDNA fragment from unintegrated BA-L DNA are presented.

L2 ANSWER 2 OF 9 COPYRIGHT 1993 ACS
AN CA116(6):46279q
TI Non-infectious HIV-1 particles and their use as vaccines
SO PCT Int. Appl., 59 pp.
AU Young, Richard A.; Baltimore, David; Aldovini, Anna; Trono, Didier; Feinberg, Mark B.
AI WO 90-US5932 16 Oct 1990
PI WO 9105860 A1 2 May 1991
PY 1991
AB Noninfectious HIV-1 particles are produced using plasmids which encode HIV-1 mutants which are defective in viral packaging. These particles may be used as vaccines. Plasmids encoding HIV-1 with a deletion in the .vphi. site and/or substitution mutations in the metal-binding motifs of the gag gene were prepd. and the constructs were introduced into COS-1 cells. HIV-1 particles were produced but the particles were not infectious (as detd. by failure to infect H9 T leukemia cell line).

L2 ANSWER 3 OF 9 COPYRIGHT 1993 ACS
AN CA115(25):272696m
TI Molecular clones of HIV-1 and their uses
SO U. S. Pat. Appl., 61 pp. Avail. NTIS Order No. PAT-APPL-6-599 491.
AU Reitz, Marvin
AI US 91-599491 31 Jan 1991
PI US 599491 A0 1 Aug 1991
PY 1991
AB The cDNA sequences representing the complete genomes of HIV-1 strains MN-PH1 and MN-ST1 are presented as in the cDNA for the env gene of a third HIV-1 strain, BA-L. The cDNAs can be used to produce anti-HIV-1 vaccines and for diagnosis of HIV-1 infection (no data). Expression plasmids for the env gene proteins of the strains were prepd. A eukaryotic expression plasmid contg. the entire MN-ST1 cDNA was prepd. for use in prodn. of the virus.

L2 ANSWER 4 OF 9 COPYRIGHT 1993 ACS
AN CA114(17):162208y
TI Production of a nonfunctional nef protein in human immunodeficiency virus type 1-infected CEM cells
SO J. Gen. Virol., 71(10), 2273-81
AU Laurent, Anne G.; Hovanessian, Ara G.; Riviere, Yves; Krust, Bernard; Regnault, Armelle; Montagnier, Luc; Findeli, Annie; Kieny, Marie Paule; Guy, Bruno
PY 1990
AB The nef gene product of the human immunodeficiency virus (HIV) is suggested to be a neg. factor involved in down-regulating viral expression by a mechanism in which the correct conformation of the nef protein is essential. The nef protein expressed by vaccinia virus recombinants is phosphorylated by protein kinase C. The present study investigated the synthesis of the nef protein and its state of phosphorylation during HIV-1 infection of a T4 cell line (CEM cells). Max. synthesis of viral proteins occurred 3 days after infection, when more than 90% of cells were producing viral proteins. The synthesis of the nef protein was detected in parallel with the env and gag proteins. As expected, the nef protein was myristylated but not phosphorylated and its half-life was less than 1 h. By the use of the polymerase chain reaction technique, the nef gene of this HIV-1 stock was isolated and sequenced. Two significant mutations were obsd. Firstly threonine, at amino acid no. 15, the site of phosphorylation by protein kinase C, was mutated into an alanine, and secondly aspartic acid of the tetrapeptide WRFD, which is probably involved in GTP binding, was mutated into an asparagine. The mutated nef gene was expressed in a vaccinia virus system, in which is was not phosphorylated and its half-life was dramatically reduced compared to the wild-type nef gene product. Furthermore, down-regulation of CD4 cell surface expression was no longer affected by the mutated nef gene. These results emphasize that phosphorylation of the nef protein provides an efficient test to monitor its biol. activity.

L2 ANSWER 5 OF 9 COPYRIGHT 1993 ACS
AN CA111(19):168198e
TI Biological and molecular characterization of human immunodeficiency virus (HIV-1BR) from the brain of a patient with progressive dementia
SO Virology, 168(1), 79-89
AU Anand, Rita; Thayer, Richard; Srinivasan, A.; Nayyar, S.; Gardner, Murray; Luciw, Paul; Dandekar, Satya
PY 1989
AB HIV-1BR was isolated from the autopsied brain tissue of a 57-yr-old man who died of progressive dementing illness. This virus was shown

to be HIV-1 by [redacted]bridization to HIV-specif[redacted] DNA probes. The expression of viral proteins as tested by radioimmunopptn. assay revealed the presence of HIV-1 specific proteins. HIV-1BR replicated in cultures of CD4+ T-lymphoid cells and induced cytopathic effects in these cells. HIV-1BR also replicated in monocytoid cell lines. The genetic nature of this isolate was detd. by mol. cloning and sequencing of the 3′-half of the genome. DNA sequence information established that HIV-1BR is a unique HIV-1 isolate. A stretch of .apprx.30 bases in the nef gene of HIV-1BR was found duplicated when compared with the other sequenced HIV-1 genomes. The functional significance of this duplication remains to be detd.

L2 ANSWER 6 OF 9 COPYRIGHT 1993 ACS
AN CA108(1):1299q
TI Complete nucleotide sequences of functional clones of the AIDS virus
SO AIDS Res. Hum. Retroviruses, 3(1), 57-69
AU Ratner, Lee; Fisher, Amanda; Jagodzinski, Linda L.; Mitsuya, Hiroaki; Liou, Ruey Shyan; Gallo, Robert C.; Wong-Staal, Flossie
PY 1987
AB To examine the mechanism of lymphocytotoxicity induced by human T-lymphotropic virus type III/lymphadenopathy assocd. virus (HTLV-III/LAV), an in vitro model has been developed. Introduction of an HTLV-III/LAV proviral clone, HXB2, into normal lymphocytes results in the prodn. of virions and cell death. The complete nucleotide sequence of the proviral form of HXB2 has now been detd. Its structure is quite similar to that previously detd. for HTLV-III/LAV clones whose biol. capacities had not previously been demonstrated. The biol. function of 2 addnl. clones of HTLV-III/LAV, BH10 and HXB3, are reported. Clone BH10 which lacks the 5′long terminal repeat sequences (LTR) and a portion of the 3′LTR is reconstituted by substituting the corresponding sequences of HXB2 and is capable of generating infectious cytopathic virions. Clone HXB3, which has been partially sequenced, is also capable of producing lymphocytopathic virus. Clone HXB3 differs from HXB2 in its lack of a termination codon in 3′orf, demonstrating that 3′orf plays no major role in virus replication or cytopathic activity. These data provide the necessary background to allow the identification of viral determinants of replication, cytopathic activity, and antigenicity using these functional proviral clones.

L2 ANSWER 7 OF 9 COPYRIGHT 1993 ACS
AN CA105(1):1450v
TI Three novel genes of human T-lymphotropic virus type III: immune reactivity of their products with sera from acquired immune deficiency syndrome patients
SO Proc. Natl. Acad. Sci. U. S. A., 83(7), 2209-13
AU Arya, Suresh K.; Gallo, Robert C.
PY 1986
AB Human T-lymphotropic virus type III or lymphoadenopathy assocd. virus (HTLV-III/LAV) is the cause of acquired immune deficiency syndrome (AIDS). In addn. to the conventional retroviral genes involved in virus replication, namely, gag, pol, and env genes, DNA sequence anal. of HTLV-III genome predicted 2 addnl. open reading frames, termed short open reading frame (sor) and 3′ open reading frame (3′ orf). Further, functional anal. revealed another gene with transactivating function, termed tat. These HTLV-III specific genes were structurally identified and functionally characterized by cDNA cloning. DNA sequence anal. of the clones shows that the tat and 3′ orf genes contain 3 exons and their transcription into functional mRNA involves 2 splicing events and that the sor gene contains .gtoreq.2 exons. In vitro transcription and translation of the cloned spliced sequences show that the sor, tat, and 3′ orf genes

code for polypeptides with apparent mobility of 24-25 kilodaltons (kDa), 14-15 kDa, and 26-28 kDa, resp. All 3 polypeptides are immune reactive and are immunogenic in the natural host. Thus, the 3 extra open reading frames of HTLV-III, 2 of which are unique to HTLV-III, are genes that function in vivo and code for 3 new and previously unrecognized HTLV-III antigens with differential immunogenicity in individuals with acquired immune deficiency syndrome and related disorders.

L2 ANSWER 8 OF 9 COPYRIGHT 1993 ACS
AN CA102(21):179952m
TI Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus
SO Nature (London), 313(6002), 450-8
AU Muesing, Mark A.; Smith, Douglas H.; Cabradilla, Cirilo D.; Benton, Charles V.; Lasky, Laurence A.; Capon, Daniel J.
PY 1985
AB The 9213-nucleotide structure of the acquired immune deficiency syndrome (AIDS)/lymphadenopathy virus has been detd. from mol. clones representing the integrated provirus and viral RNA. The sequence reveals that the virus is highly polymorphic and lacks significant nucleotide homol. with type C retroviruses characterized previously. Together with an anal. of the 2 major viral subgenomic RNAs, these studies establish the coding frames for the gag, pol and env genes and predict the expression of a novel gene at the 3′ end of the genome unrelated to the X genes of human T-lymphotrophic virus I and II.

L2 ANSWER 9 OF 9 COPYRIGHT 1993 ACS
AN CA102(15):126416h
TI Nucleotide sequence of the AIDS virus, LAV
SO Cell (Cambridge, Mass.), 40(1), 9-17
AU Wain-Hobson, Simon; Sonigo, Pierre; Danos, Olivier; Cole, Stewart; Alizon, Marc
PY 1985
AB The complete 9193-nucleotide sequence of the probable causative agent of acquired immune deficiency syndrome (AIDS), lymphadenopathy-assocd. virus (LAV), was detd. The deduced genetic structure is unique; it shows, in addn. to the retroviral gag, pol, and env genes, 2 novel open reading frames which were designated Q and F. Remarkably, Q is located between pol and env, and F is half-encoded by the U3 element of the long terminal repeat. Thus, LAV is distinct from the previously characterized family of human T cell leukemia (lymphoma) viruses.

E1 THROUGH E14 ASSIGNED

=> fil reg; s e1-e14
FILE 'REGISTRY' ENTERED AT 14:43:18 ON 26 APR 93
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 1993 American Chemical Society (ACS)

STRUCTURE FILE UPDATES: 23 APR 93 HIGHEST RN 147199-92-6
DICTIONARY FILE UPDATES: 25 APR 93 HIGHEST RN 147199-92-6

1 137574-23-3/RN
1 102686-56-6/RN
1 111804-75-2/RN
1 111804-83-2/RN
1 123056-88-2/RN

```
                1 133172-96-0/RN
                1 138362-52-4/RN
                1 138362-53-5/RN
                1 138362-54-6/RN
                1 138362-55-7/RN
                1 138362-56-8/RN
                1 95568-14-2/RN
                1 96098-36-1/RN
                1 96098-41-8/RN
L3             14 (137574-23-3/RN OR 102686-56-6/RN OR 111804-75-2/RN OR 111
                  804-83-2/RN OR 123056-88-2/RN OR 133172-96-0/RN OR 138362-
                  52-4/RN OR 138362-53-5/RN OR 138362-54-6/RN OR 138362-55-7
                  /RN OR 138362-56-8/RN OR 95568-14-2/RN OR 96098-36-1/RN OR
                   96098-41-8/RN)

=> d 1-14 .bevreg; fil ca; e alizon, m/au 10

L3   ANSWER 1 OF 14  COPYRIGHT 1993 ACS
RN   138362-56-8  REGISTRY
CN   Deoxyribonucleic acid (human immunodeficiency provirus 1 clone
     pA14-15HXB) (9CI)  (CA INDEX NAME)
SQL  9609
MF   Unspecified
CI   MAN

L3   ANSWER 2 OF 14  COPYRIGHT 1993 ACS
RN   138362-55-7  REGISTRY
CN   Deoxyribonucleic acid (human immunodeficiency provirus 1 clone
     pA15HXB) (9CI)  (CA INDEX NAME)
SQL  9606
MF   Unspecified
CI   MAN

L3   ANSWER 3 OF 14  COPYRIGHT 1993 ACS
RN   138362-54-6  REGISTRY
CN   Deoxyribonucleic acid (human immunodeficiency provirus 1 clone
     pA4HXB) (9CI)  (CA INDEX NAME)
SQL  9606
MF   Unspecified
CI   MAN

L3   ANSWER 4 OF 14  COPYRIGHT 1993 ACS
RN   138362-53-5  REGISTRY
CN   Deoxyribonucleic acid (human immunodeficiency provirus 1 clone
     pA3HXB) (9CI)  (CA INDEX NAME)
SQL  9607
MF   Unspecified
CI   MAN

L3   ANSWER 5 OF 14  COPYRIGHT 1993 ACS
RN   138362-52-4  REGISTRY
CN   Deoxyribonucleic acid (human immunodeficiency provirus 1 clone
     bCA20-W13) (9CI)  (CA INDEX NAME)
SQL  9613
MF   Unspecified
CI   MAN

L3   ANSWER 6 OF 14  COPYRIGHT 1993 ACS
RN   137574-23-3  REGISTRY
CN   Deoxyribonucleic acid (human immunodeficiency provirus 1 clone
     .lambda.BA-L1 gene env plus 5'- and 3'-flanking region fragment)
     (9CI)  (CA INDEX NAME)
```

SQL 3807
MF Unspecified
CI MAN

L3 ANSWER 7 OF 14 COPYRIGHT 1993 ACS
RN 133172-96-0 REGISTRY
CN Deoxyribonucleic acid (human immunodeficiency provirus 1 gene nef') (9CI) (CA INDEX NAME)
SQL 621
MF Unspecified
CI MAN

L3 ANSWER 8 OF 14 COPYRIGHT 1993 ACS
RN 123056-88-2 REGISTRY
CN Deoxyribonucleic acid (human immunodeficiency provirus clone pATZ6 gene nef) (9CI) (CA INDEX NAME)
SQL 657
MF Unspecified
CI MAN

L3 ANSWER 9 OF 14 COPYRIGHT 1993 ACS
RN 111804-83-2 REGISTRY
CN Deoxyribonucleic acid (human immunodeficiency provirus clone HXB2 13-kilodalton protein gene) (9CI) (CA INDEX NAME)
SQL 621
MF Unspecified
CI MAN

L3 ANSWER 10 OF 14 COPYRIGHT 1993 ACS
RN 111804-75-2 REGISTRY
CN Deoxyribonucleic acid (human immunodeficiency provirus clone HXB2) (9CI) (CA INDEX NAME)
SQL 9177
MF Unspecified
CI MAN

L3 ANSWER 11 OF 14 COPYRIGHT 1993 ACS
RN 102686-56-6 REGISTRY
CN Deoxyribonucleic acid (human immunodeficiency provirus clone pSP-12 27-kilodalton protein gene) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN Deoxyribonucleic acid (human T-cell leukemia provirus type III clone pSP-12 27-kilodalton protein gene)
SQL 642
MF Unspecified
CI MAN

L3 ANSWER 12 OF 14 COPYRIGHT 1993 ACS
RN 96098-41-8 REGISTRY
CN Deoxyribonucleic acid (human immunodeficiency provirus clone H9pv.22 protein E' gene) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN Deoxyribonucleic acid (lymphadenopathy/AIDS provirus clone H9pv.22 protein E' gene)
SQL 621
MF Unspecified
CI MAN

L3 ANSWER 13 OF 14 COPYRIGHT 1993 ACS
RN 96098-36-1 REGISTRY
CN Deoxyribonucleic acid (human immunodeficiency provirus clone H9pv.22) (9CI) (CA INDEX NAME)

```
OTHER NAMES:
CN   Deoxyribonucleic acid (lymphadenopathy/AIDS provirus clone H9pv.22)
SQL  9213
MF   Unspecified
CI   MAN

L3   ANSWER 14 OF 14  COPYRIGHT 1993 ACS
RN   95568-14-2  REGISTRY
CN   Deoxyribonucleic acid (human immunodeficiency provirus clone
     .lambda.J19) (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   Deoxyribonucleic acid (lymphadenopathy-associated provirus clone
     .lambda.J19)
SQL  9193
MF   Unspecified
CI   MAN


FILE 'CA' ENTERED AT 14:43:48 ON 26 APR 93
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 1993 AMERICAN CHEMICAL SOCIETY (ACS)

FILE COVERS 1967 - 13 Apr 93 (930413/ED)  VOL 118 ISS 16.
For OFFLINE Prints or Displays, use the ABS or ALL formats to obtain
abstract graphic structures.  The AB format DOES NOT display structure
diagrams.


E1          22     ALIZON, J/AU
E2           3     ALIZON, JOSEPH/AU
E3           2 --> ALIZON, M/AU                    - Author(s)
E4          25     ALIZON, MARC/AU
E5           3     ALJ, A/AU
E6           1     ALJ, A E/AU
E7           1     ALJABAB, A/AU
E8           1     ALJABRE, S H M/AU
E9           8     ALJADEFF, GLADIS/AU
E10          1     ALJADEHEFF, GLADIS/AU

=> s e3-e4; e sonico, p/au 10
             2 "ALIZON, M"/AU
            25 "ALIZON, MARC"/AU
L4          27 ("ALIZON, M"/AU OR "ALIZON, MARC"/AU)


E1           6     SONICH, V P/AU
E2           1     SONICH, V V/AU
E3           0 --> SONICO, P/AU
E4           2     SONIDIS, GEORGE P/AU
E5           1     SONIDO, E P/AU
E6           4     SONIE, K C/AU
E7           3     SONIER, FELIX/AU
E8           1     SONIER, FERNAND/AU
E9           3     SONIGO, P/AU
E10         26     SONIGO, PIERRE/AU

=> e stewart, c/au 10
E1           2     STEWART, BRUCE N/AU
E2           1     STEWART, BURCH BYRON/AU
E3          18 --> STEWART, C/AU
E4          14     STEWART, C A/AU
E5           1     STEWART, C A JR/AU
```

```
E6            5      STEWART, C B/AU
E7           10      STEWART, C C/AU
E8            7      STEWART, C D/AU
E9            1      STEWART, C E/AU
E10           6      STEWART, C E E/AU

=> s e3; e stewart, cole/au 10
L5           18 "STEWART, C"/AU


E1            2      STEWART, CLIVE EDWARD E/AU
E2            1      STEWART, CLIVE EDWARD ERNEST/AU
E3            1 --> STEWART, COLE/AU
E4            3      STEWART, COLIN/AU
E5            1      STEWART, COLIN C/AU
E6            2      STEWART, COLIN CROSBIE/AU
E7           11      STEWART, COLIN L/AU
E8           15      STEWART, COLIN S/AU
E9            1      STEWART, COLIN SAMUEL/AU
E10           4      STEWART, CONSTANCE B/AU

=> s e3; s l5 or l6; e danos, o/au 10
L6            1 "STEWART, COLE"/AU


L7           19 L5 OR L6


E1           57      DANOS, MICHAEL/AU
E2            1      DANOS, MICHEL/AU
E3            6 --> DANOS, O/AU
E4            1      DANOS, OLIVER/AU
E5           22      DANOS, OLIVIER/AU
E6            1      DANOS, OLIVIER F/AU
E7            1      DANOS, P T/AU
E8            1      DANOS, R J/AU
E9            5      DANOS, ROBERT J/AU
E10           1      DANOS, SAVAS C/AU

=> s e3-e6; e wain-hobson, s/au 9
              6 "DANOS, O"/AU
              1 "DANOS, OLIVER"/AU
             22 "DANOS, OLIVIER"/AU
              1 "DANOS, OLIVIER F"/AU
L8           30 ("DANOS, O"/AU OR "DANOS, OLIVER"/AU OR "DANOS, OLIVIER"/A
                U OR "DANOS, OLIVIER F"/AU)


E1            3      WAIN, WILLIAM H/AU
E2            1      WAIN, WILLIAM HENRY/AU
E3            0 --> WAIN-HOBSON, S/AU
E4            2      WAINAI, HIDEKI/AU
E5            1      WAINAI, TASUKU/AU
E6            1      WAINAI, TOHORU/AU
E7           28      WAINAI, TOHRU/AU
E8            1      WAINAI, TOORU/AU
E9            2      WAINAI, TORU/AU

=> e wain, s/au 10
E1            1      WAIN, RUSSELL/AU
E2            1      WAIN, RUSSELL EDMUND/AU
E3            0 --> WAIN, S/AU
```

```
E4            1      WA[redacted], W/AU
E5            8      WAIN, W H/AU
E6            3      WAIN, WILLIAM H/AU
E7            1      WAIN, WILLIAM HENRY/AU
E8            2      WAINAI, HIDEKI/AU
E9            1      WAINAI, TASUKU/AU
E10           1      WAINAI, TOHORU/AU

=> e hobson, s/au 10
E1            1      HOBSON, ROY BAXTER/AU
E2            1      HOBSON, RUSSELL B JR/AU
E3            4 --> HOBSON, S/AU
E4            1      HOBSON, SIMON WAIN/AU
E5            6      HOBSON, T/AU
E6            4      HOBSON, W/AU
E7            8      HOBSON, W C/AU
E8          115      HOBSON, W S/AU
E9            2      HOBSON, W T/AU
E10           8      HOBSON, WILLIAM/AU

=> s e3-e4
              4 "HOBSON, S"/AU
              1 "HOBSON, SIMON WAIN"/AU
L9            5 ("HOBSON, S"/AU OR "HOBSON, SIMON WAIN"/AU)

=> s l4 and l7 and l8 and l9; s l4 and (l7 or l8 or l9); s l7 and (l8 or l
9); s l8 and l9
L10           0 L4 AND L7 AND L8 AND L9


L11           3 L4 AND (L7 OR L8 OR L9)


L12           0 L7 AND (L8 OR L9)


L13           0 L8 AND L9

=> s (l4 or l7 or l8 or l9) and (lav or lymphadenopath? or htlv or hiv or
lymphotrop? or human(2w)virus?)/ab,bi
            245 LAV/AB
             98 LAV/BI
            264 LYMPHADENOPATH?/AB
            130 LYMPHADENOPATH?/BI
           1504 HTLV/AB
            827 HTLV/BI
           6282 HIV/AB
           5288 HIV/BI
            667 LYMPHOTROP?/AB
            536 LYMPHOTROP?/BI
         307992 HUMAN/AB
         309481 HUMAN/BI
          84015 VIRUS?/AB
         123349 VIRUS?/BI
           8477 HUMAN(2W)VIRUS?
L14          30 (L4 OR L7 OR L8 OR L9) AND (LAV OR LYMPHADENOPATH? OR HTLV
                OR HIV OR LYMPHOTROP? OR HUMAN(2W)VIRUS?)/AB,BI

=> s l14 and clon?/ab,bi
          84117 CLON?/AB
          54670 CLON?/BI
L15          17 L14 AND CLON?/AB,BI
```

```
=> s l15 and sequenc?/ab,bi
        199071 SEQUENC?/AB
        103235 SEQUENC?/BI
L16          14 L15 AND SEQUENC?/AB,BI

=> s (l11 or l16) not l2
L17          15 (L11 OR L16) NOT L2

=> d 1-15 .beverly; fil biosi; s alizon m ?/au; s sonico p ?/au; s stewart
 c ?/au; s danos o ?/au; s (hobson s ? or wain s ?)/au

L17  ANSWER 1 OF 15  COPYRIGHT 1993 ACS
AN   CA116(5):39665j
TI   Immunogenic peptides of a variant of LAV (
   lymphadenopathy virus)
SO   U.S., 49 pp.
AU   Alizon, Marc; Sonigo, Pierre; Wain-Hobson, Simon; Montagnier, Luc
AI   US 87-38332  13 Apr 1987
PI   US 5034511  A  23 Jul 1991
PY   1991
AB   Immunogenic peptide sequences from LAVELI are presented.
     An immunogenic compn. comprising such a peptide and a physiol.
     acceptable carrier as well as a diagnostic kit for detecting
     antibodies to LAV comprising such a peptide and a reagent
     for detecting the formation of peptide/antibody complex are also
     claimed. Sequences are claimed from env, gag, and pol
     proteins. The complete cDNA of LAVELI is presented. The
   sequence was compared with those for other LAV.

L17  ANSWER 2 OF 15  COPYRIGHT 1993 ACS
AN   CA112(1):2059f
TI   Expression vectors for manufacture of human
     immunodeficiency virus 2 (HIV2) proteins
SO   Fr. Demande, 31 pp.
AU   Kieny, Marie Paule; Rautmann, Guy; Guy, Bruno; Montagnier, Luc;
     Alizon, Marc; Girard, Marc
AI   FR 87-12396  7 Sep 1987
PI   FR 2620030  A1  10 Mar 1989
PY   1989
AB   Viral or plasmid vectors which can be used to manuf. HIV2 proteins
     in eukaryotes or prokaryotes are described. The HIV2 proteins can be
     used as vaccines or to prep. antibodies. Both proteins and
     antibodies can be used in diagnosis. The cDNA for HIV2 protein F was
   cloned in plasmid pTG186POLY, and this plasmid used to prep.
     recombinant vaccinia virus by std. means. BHK21 cells were infected
     with this recombinant virus. Protein which was recognized by serum
     from HIV2 pos. patients was produced by these transformants.

L17  ANSWER 3 OF 15  COPYRIGHT 1993 ACS
AN   CA111(1):2164r
TI   Peptides having immunological properties of HIV-2 (
   human immunodeficiency virus) for diagnosis and
     vaccines and simian immunodeficiency virus genome cDNA
   sequence
SO   PCT Int. Appl., 96 pp.
AU   Alizon, Marc; Montagnier, Luc; Guetard, Denise; Clavel, Francois;
     Sonigo, Pierre; Guyader, Mireille; Tiollais, Pierre; Chakrabarti,
     Lisa; Desrosiers, Ronald
AI   WO 88-FR25  15 Jan 1988
PI   WO 8805440  A1  28 Jul 1988
PY   1988
```

AB Peptides having immunol. properties in com with HIV-2, particularly the envelope glycoprotein of HIV-21, and with the glycoprotein of SIV-1 (simian immunodeficiency virus) are useful in detecting infection with HIV-2 and in vaccines. Diagnostic kits and cDNA sequences esp. for SIV-1 macaque are also included. The DNA of HUT 78 cells infected with SIV-1 of macaque was partially digested with restriction endonuclease Sau 345 and cloned in the BamHI of .lambda. to construct a gene bank. The recombinant phages were screened using sequences of HIV-2. One clone, .lambda.SIV-1, had a 16.5-kilobase insert comprising the entire provirus genome lacking only 250 bases at the left long terminal repeat region. The nucleotide sequence was detd. by the dideoxynucleotide method after subcloning in phage M13mp8.

L17 ANSWER 4 OF 15 COPYRIGHT 1993 ACS
AN CA110(17):152651r
TI Envelope antigens of lymphadenopathy-associated virus and their applications
SO PCT Int. Appl., 78 pp.
AU Montagnier, Luc; Krust, Bernard; Chamaret, Solange; Clavel, Francois; Chermann, Jean Claude; Barre-sinoussi, Francoise; Alizon, Marc; Sonigo, Pierre; Stewart, Cole; et al.
AI WO 85-EP548 18 Oct 1985
PI WO 8602383 A1 24 Apr 1986
PY 1986
AB Purified expression products of DNA sequences derived from the lymphadenopathy-assocd. virus (LAV) genome, particularly a 110,000-mol.-wt. glycoprotein or derived antigenic peptides which are recognized by human sera contg. antibodies against LAV, are prepd. The glycoprotein is used in the prepn. of monoclonal antibodies and in the prodn. of an immunogenic compn. capable of neutralizing LAV. The glycoprotein or polypeptides are also useful in the diagnosis of LAV antibodies in sera of patients. T-lymphocytes derived from healthy and LAV1-infected donors were cultivated in a nondenaturing medium contg. cysteine-35S. The supernatant from the culture medium was centrifuged at 10,000 rpm for 10 min to remove the nonviral components, then at 45,000 rpm for 20 min to sediment the virus. The virus pellet was then lysed by detergent in the presence of aprotinin and the envelope glycoprotein (gp110) was purified by affinity chromatog. on Sephrose-Con A and eluted with O-methyl-.alpha.-D-mannopyranoside. The gp110 was used to immunize mice for the prodn. of monoclonal antibodies by std. hybridoma methodol. The sequencing and detn. of peptide or protein sites of particular interest were carried out on a recombinant phage corresponding to .lambda.J19 or LAV-Ia.

L17 ANSWER 5 OF 15 COPYRIGHT 1993 ACS
AN CA109(15):123790j
TI Variants of lymphadenopathy-associated viruses, their cDNA and protein sequences and their use, particularly for diagnostic purposes and for the preparation of immunogenic compositions
SO PCT Int. Appl., 72 pp.
AU Alizon, Marc; Sonigo, Pierre; Wain-Hobson, Simon; Montagnier, Luc
AI WO 87-EP326 22 Jun 1987
PI WO 8707906 A1 30 Dec 1987
PY 1987
AB Two new variants of lymphadenopathy-assocd. viruses (LAV) designated LAVILI and LAVMAL are isolated and their genomes characterized. Their DNAs and antigens can be used for the

diagnosis of AIDS and prodn. of vaccines against AIDS. The viruses were isolated from African patients from Zaire. The genetic organization of the two new isolates, esp. the region between the pol and env genes, is identical to that of the other isolates. The sizes of the U3, R, and U5 elements of the long terminal repeat are also conserved. Substantial differences are obsd. in the primary structure of their proteins; the envelope is more variable that the gag and pol gene proteins.

L17 ANSWER 6 OF 15 COPYRIGHT 1993 ACS
AN CA109(11):89337e
TI Retrovirus of the human immunodeficiency virus 2 (HIV-2) type capable of inducing AIDS, its antigenic and nucleic acid constituents, and diagnostic and therapeutic methods and kits
SO PCT Int. Appl., 117 pp.
AU Montagnier, Luc; Chamaret, Solange; Guetard, Denise; Alizon, Marc; Clavel, Francois; Guyader, Mireille; Sonigo, Pierre; Brun-Vezinet, Francoise; Rey, Marianne; et al.
AI WO 87-FR25 22 Jan 1987
PI WO 8704459 A1 30 Jul 1987
PY 1987
AB Retrovirus HIV-2 and its antigenic and nucleic acid components are useful in diagnostic (e.g. antibody immunoassays) and therapeutic methods and kits. Protein antigens p12, p16, p26, and gp140 and genetic material have been prepd. Glycoprotein gp140 is particularly useful in immunogenic compns. Nucleotide sequences useful as hybridization probes are disclosed. HIV of patients from west Africa was isolated by stimulating their peripheral blood lymphocytes (PBLs) with PHA and cultivating in coculture with normal PBLs so stimulated and maintained in the presence of interleukin-2. The viruses were centrifuged, lysed, and deposited on nitrocellulose. The samples were treated with an HIV-1 probe corresponding to the complete genome of LAVBRU or an HIV-2 probe derived from a 2-kb cDNA clone of LAV-2ROD, both labeled with 32P, under stringent hybridization conditions. All of the virus samples hybridized with the HIV-2 probe only.

L17 ANSWER 7 OF 15 COPYRIGHT 1993 ACS
AN CA108(23):199491n
TI Preparation of recombinant viral vectors encoding human immunodeficiency virus (HIV) glycoprotein for use as anti-AIDS vaccine
SO Fr. Demande, 36 pp.
AU Kieny, Marie Paule; Rautmann, Guy; Lecocq, Jean Pierre; Hobson, Simon Wain; Girard, Marc; Montagnier, Luc
AI FR 86-5043 8 Apr 1986
PI FR 2596771 A1 9 Oct 1987
PY 1987
AB Viral vectors which encode HIV env protein or variants thereof are constructed, mammalian cells are infected with them, and the immunogenicity of the recombinant proteins are analyzed. Plasmid pTG1125 contg., inserted into the vaccinia virus thymidine kinase gene, the HIV env gene under the control of the vaccinia virus 7.5K protein gene promoter was constructed. Viral vector VV.TG. eLAV 1125 was prepd. by in vivo recombination of pTG1125 with vaccinia virus. BHK21 cells infected with this vector produced glycoproteins of mol. wt. 160, 120, and 40 kilodaltons which were recognized by antiserum isolated from AIDS patients. Balb/c mice infected with this vector produced antibodies which reacted with 160- and 40-kilodalton proteins in sera of AIDS patients.

L17 ANSWER 8 OF 15 COPYRIGHT 1993 ACS
AN CA108(13):107210u
TI Sequence analysis of the human immune deficiency virus type 2
SO UCLA Symp. Mol. Cell. Biol., New Ser., 71(Hum. Retroviruses, Cancer, AIDS), 31-42
AU Guyader, M.; Emerman, M.; Sonigo, P.; Clavel, F.; Montagnier, L.; Alizon, M.
PY 1988
AB Cloned cDNA probes made from human immunodeficiency type 2 virus (HIV-2) were used to screen a genomic library made from a T4 cell line infected with the ROD isolate of HIV-2. Lambda clones contg. proviral DNA were characterized by restriction mapping, and then used to det. the complete 9671-nucleotide sequence of the genome. The genomic organization of HIV-2 was 5'LTR-gag-pol-central region-env-orfF-3'LTR; the central region contained 4 genes related to those of HIV-1 (sor, R, tat, and art) as well as a 5th gene (designated X) with no counterpart in HIV-1. HIV-1 and HIV-2 differed significantly in terms of nucleotide and amino acid sequence. The more conserved gag and pol genes displayed only 56 and 60% nucleotide sequence homol. and both <60% of amino acid identity. Calcn. of the nucleotide sequence homol. for the other genes gave even lower values, giving HIV-1 and 2 overall 42% homologous. To det. whether or not the tat gene of HIV-1 could trans-activate the LTR of HIV-2 and vice versa, SW480 cells were cotransfected with subgenomic fragments of HIV-1 or HIV-2 and pHIV2-CAT or a plasmid pHIV1-CAT which contained U3-R of HIV-1. Both HIV-1 and HIV-2 LTRs were substantially activated by the HIV-1 tat gene.

L17 ANSWER 9 OF 15 COPYRIGHT 1993 ACS
AN CA108(1):1300h
TI Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses
SO Nature (London), 328(6130), 543-7
AU Chakrabarti, Lisa; Guyader, Mireille; Alizon, Marc; Daniel, Muthiah D.; Desrosiers, Ronald C.; Tiollais, Pierre; Sonigo, Pierre
PY 1987
AB The complete genome of the proviral form of simian immunodeficiency virus isolated from a naturally infected macaque was cloned (.lambda.SIV1) and sequenced. The genome of SIVmac was 9643 nucleotides long with its open reading frames and was organized (5'LTR-gag-pol-central region-env-F-3'LTR) in a manner typical of a lentivirus. Comparisons of the proteins of SIV with those of HIV-1 and HIV-2 quantified the relatedness of these viruses.

L17 ANSWER 10 OF 15 COPYRIGHT 1993 ACS
AN CA106(11):79452n
TI Molecular cloning and polymorphism of the human immune deficiency virus type 2
SO Nature (London), 324(6098), 691-5
AU Clavel, Francois; Guyader, Mireille; Guetard, Denise; Salle, Mireille; Montagnier, Luc; Alizon, Marc
PY 1986
AB A novel retrovirus, human immune deficiency virus type 2 (HIV-2), was isolated and characterized. Hybridization expts. indicated that there are substantial

differences between the DNA sequences of HIV-2 and HIV-1. Moreover, the serol. cross-reactivity of the proteins of the 2 viruses is restricted to the core protein. The 9.5-kilobase genome of HIV-2 was cloned. Different isolates of HIV-2 exhibited restriction site polymorphism in their DNAs. The relationship of HIV-2 with other human and simian retroviruses is discussed.

L17 ANSWER 11 OF 15 COPYRIGHT 1993 ACS
AN CA106(5):28512z
TI Cloned DNA sequences, hybridizable with genomic RNA of lymphadenopathy-associated virus (lav)
SO PCT Int. Appl., 39 pp.
AU Alizon, Marc; Barre Sinoussi, Francoise; Sonigo, Pierre; Tiollais, Pierre; Chermann, Jean Claude; Montagnier, Luc; Wain-Hobson, Simon
AI WO 85-EP487 18 Sep 1985
PI WO 8601827 A1 27 Mar 1986
PY 1986
AB Cloned DNA fragments contg. sequences hybridizable to genomic RNA and DNA of lymphadenopathy -assocd. retrovirus (LAV) are obtained from a cDNA library of the LAV genome. These DNA fragments are useful as hybridization probes for detection of LAV in biol. samples taken from persons possibly afflicted with AIDS. The complete sequence and restriction map of the LAV provirus genome are presented.

L17 ANSWER 12 OF 15 COPYRIGHT 1993 ACS
AN CA105(21):185219f
TI AIDS virus env protein expressed from a recombinant vaccinia virus
SO Bio/Technology, 4(9), 790-5
AU Kieny, M. P.; Rautmann, G.; Schmitt, D.; Dott, K.; Wain-Hobson, S.; Alizon, M.; Girard, M.; Chamaret, S.; Laurent, A.; et al.
PY 1986
AB Lymphadenopathy-assocd. virus (LAV) in the causative agent of AIDS, the acquired immunodeficiency syndrome. A retrovirus of the lentivirus group, LAV carries a single major target antigen at its surface: the env protein. The env coding sequence was introduced into a vaccinia virus vector. The live recombinant virus, VVTGeLAV, dets. the prodn. of env protein in infected mammalian cells. The recombinant protein reacts with sera from AIDS patients and appear to be processed and glycosylated in a manner identical to authentic env of LAV retrovirus. Inoculation of mice with VVTGeLAV elicits high titers of antisera recognizing vaccinia determinants but only low titers of antibody recognizing env proteins of LAV. Cells infected with the recombinant virus rapidly liberate a processed form of the env protein into the culture medium. This shedding of surface antigen from AIDS virus may play a role in the pathophysiol. of the disease.

L17 ANSWER 13 OF 15 COPYRIGHT 1993 ACS
AN CA105(9):73424n
TI Lymphadenopathy/AIDS virus: genetic organization and relationship to animal lentiviruses
SO Anticancer Res., 6(3, Pt. B), 403-12
AU Alizon, Marc; Montagnier, Luc
PY 1986
AB A review with 46 refs. on the mol. characterization of the probable agent of the acquired immune deficiency syndrome (AIDS), the lymphadenopathy/AIDS virus (LAV). Mol. cloning and complete nucleotide sequencing of LAV allows a detailed comparison with other AIDS virus

isolates, as well as with other human and animal retroviruses. The AIDS virus is closely related to visna virus, prototype of the lentiviruses, whereas the other human retroviruses, i.e., human T-cell leukemia viruses type I and II (HTLV-I and II), are quite remote in the evolution.

L17 ANSWER 14 OF 15 COPYRIGHT 1993 ACS
AN CA103(19):155030d
TI Nucleotide sequence of the Visna lentivirus: relationship to the AIDS virus
SO Cell (Cambridge, Mass.), 42(1), 369-82
AU Sonigo, Pierre; Alizon, Marc; Staskus, Katherine; Klatzmann, David; Cole, Stewart; Danos, Olivier; Retzel, Ernest; Tiollais, Pierre; Haase, Ashley; Wain-Hobson, Simon
PY 1985
AB The complete 9202 nucleotide sequence of the visna lentivirus was detd. The deduced genetic organization most closely resembles that of the AIDS retrovirus in that there is a novel central region sepg. pol and env. Moreover, there is a close phylogenetic relation between the conserved reverse transcriptase and endonuclease/integrase domains of the visna and AIDS viruses. These findings support the inclusion of the AIDS virus in the retroviral subfamily Lentivirinae.

L17 ANSWER 15 OF 15 COPYRIGHT 1993 ACS
AN CA102(9):73509g
TI Molecular cloning of lymphadenopathy-associated virus
SO Nature (London), 312(5996), 757-60
AU Alizon, Marc; Sonigo, Pierre; Barre-Sinoussi, Francoise; Chermann, Jean Claude; Tiollais, Pierre; Montagnier, Luc; Wain-Hobson, Simon
PY 1985
AB DNA complementary to human lymphadenopathy -assocd. virus (LAV) RNA was cloned on plasmid pBR327, and the recombinant DNA was used to transform Escherichia coli. Plasmid pLAV13 carrying a 2.5-kilobase insert was isolated and its nick-translated DNA used as a hybridization probe to detect virion RNA in infected cells. LAV virion RNA was detected in infected normal T-cells, FR8 and other B cell lines, CEM cells, and bone marrow cells from a hemophiliac with AIDS, but not in uninfected normal T lymphocyte cells or normal liver. Plasmid pLAV13, which did not integrate into the human genome, detected both RNA and integrated DNA forms in LAV-infected cells. Genomic LAV sequences were similarly cloned by inserting HindIII digests of genomic DNA of LAV-infected T cells into a phage .lambda. vector; 5 recombinants that hybridized with nick-translated pLAV13 were obtained.

FILE 'BIOSIS' ENTERED AT 14:51:16 ON 26 APR 93
COPYRIGHT (C) 1993 BIOSIS(R)

FILE COVERS 1969 TO DATE.
CAS REGISTRY NUMBERS AND CHEMICAL NAMES (CNs) PRESENT FROM JANUARY 1969 TO DATE.

RECORDS LAST ADDED: 23 April 1993 (930423/ED) BA9510 BR4409
CAS REGISTRY NUMBERS (R) LAST ADDED: 24 April 1993 (930424/UP)

L18 35 ALIZON M ?/AU

```
L19           0 SONICO P ?/AU


L20         878 STEWART C ?/AU


L21          36 DANOS O ?/AU


             17 HOBSON S ?/AU
             16 WAIN S ?/AU
L22          33 (HOBSON S ? OR WAIN S ?)/AU

=> s l18 and l20 and l21 and l22; s l18 and (l20 or l21 or l22); s l20 and
 (l21 or l22); s l21 and l22
L23           0 L18 AND L20 AND L21 AND L22


L24           3 L18 AND (L20 OR L21 OR L22)


L25           0 L20 AND (L21 OR L22)


L26           0 L21 AND L22

=> s (l18 or l20 or l21 or l22) and (lav or lymphadenopath? or htlv or hiv
 or lymphotrop? or human(2w)virus?)
            583 LAV
           4873 LYMPHADENOPATH?
           4285 HTLV
          30024 HIV
           4762 LYMPHOTROP?
        2803320 HUMAN
         240513 VIRUS?
          56380 HUMAN(2W)VIRUS?
L27          34 (L18 OR L20 OR L21 OR L22) AND (LAV OR LYMPHADENOPATH? OR
                HTLV OR HIV OR LYMPHOTROP? OR HUMAN(2W)VIRUS?)

=> s l27 and clon? and sequenc?
        128815 CLON?
        191372 SEQUENC?
L28           2 L27 AND CLON? AND SEQUENC?

=> s l24 or l28; fil medl; s l28; s l18; s l20; s l21; s l22; s l19
L29           5 L24 OR L28


FILE 'MEDLINE' ENTERED AT 14:55:32 ON 26 APR 93
COPYRIGHT (C) 1993 U.S. National Library of Medicine  (NLM)

 FILE LAST UPDATED: 23 APR 93 (930423/UP).  FILE COVERS 1972 TO DATE.
 ***FILE NOW RELOADED WITH 1993 MESH. NEW SUBHEADING PRE-EXPLOSION
 FEATURE NOW AVAILABLE. SEE HELP PRE-EXPLOSION.***

 ************************************************************
 * LIMITING SEARCH FUNCTIONS (L#/MAJ, ETC.) TEMORARILY UNAVAILABLE.*
 ******************SEE HELP LLIMIT FOR DETAILS.**********************

 SEE HELP CTAG FOR CHECK TAGS, HELP SUBHEADING FOR SUBHEADINGS AND
```

```
 HELP TREE FOR TREE NUMBER CATEGORY.

 +QLF/CT SHOWS YOU THE ALLOWABLE QUALIFIERS OF A TERM.


            18 ALIZON M ?/AU
           499 STEWART C ?/AU
            25 DANOS O ?/AU
            12 HOBSON S ?/AU
            11 WAIN S ?/AU
           880 LAV
          5422 LYMPHADENOPATH?
          6592 HTLV
         29914 HIV
          1976 LYMPHOTROP?
       4131623 HUMAN
        179524 VIRUS?
         15305 HUMAN(2W)VIRUS?
        115931 CLON?
        207238 SEQUENC?
L30          7 L27 AND CLON? AND SEQUENC?


L31         18 ALIZON M ?/AU


L32        499 STEWART C ?/AU


L33         25 DANOS O ?/AU


            12 HOBSON S ?/AU
            11 WAIN S ?/AU
L34         23 (HOBSON S ? OR WAIN S ?)/AU


L35          0 SONICO P ?/AU

=> s l31 and l32 and l33 and l34; s l31 and (l32 or l33 or l34); s l32 and
 (l33 or l34); s l33 and l34
L36          0 L31 AND L32 AND L33 AND L34


L37          2 L31 AND (L32 OR L33 OR L34)


L38          0 L32 AND (L33 OR L34)


L39          0 L33 AND L34

=> s l30 or l37
L40          7 L30 OR L37

=> dup rem l29,l40
FILE 'BIOSIS' ENTERED AT 14:57:51 ON 26 APR 93
COPYRIGHT (C) 1993 BIOSIS(R)

FILE 'MEDLINE' ENTERED AT 14:57:51 ON 26 APR 93
COPYRIGHT (C) 1993 U.S. National Library of Medicine  (NLM)
PROCESSING COMPLETED FOR L29
```

```
 HELP TREE FOR TREE NUMBER CATEGORY.

 +QLF/CT SHOWS YOU THE ALLOWABLE QUALIFIERS OF A TERM.


            18 ALIZON M ?/AU
           499 STEWART C ?/AU
            25 DANOS O ?/AU
            12 HOBSON S ?/AU
            11 WAIN S ?/AU
           880 LAV
          5422 LYMPHADENOPATH?
          6592 HTLV
         29914 HIV
          1976 LYMPHOTROP?
       4131623 HUMAN
        179524 VIRUS?
         15305 HUMAN(2W)VIRUS?
        115931 CLON?
        207238 SEQUENC?
L30          7 L27 AND CLON? AND SEQUENC?


L31         18 ALIZON M ?/AU


L32        499 STEWART C ?/AU


L33         25 DANOS O ?/AU


            12 HOBSON S ?/AU
            11 WAIN S ?/AU
L34         23 (HOBSON S ? OR WAIN S ?)/AU


L35          0 SONICO P ?/AU

=> s l31 and l32 and l33 and l34; s l31 and (l32 or l33 or l34); s l32 and
 (l33 or l34); s l33 and l34
L36          0 L31 AND L32 AND L33 AND L34


L37          2 L31 AND (L32 OR L33 OR L34)


L38          0 L32 AND (L33 OR L34)


L39          0 L33 AND L34

=> s l30 or l37
L40          7 L30 OR L37

=> dup rem l29,l40
FILE 'BIOSIS' ENTERED AT 14:57:51 ON 26 APR 93
COPYRIGHT (C) 1993 BIOSIS(R)

FILE 'MEDLINE' ENTERED AT 14:57:51 ON 26 APR 93
COPYRIGHT (C) 1993 U.S. National Library of Medicine  (NLM)
PROCESSING COMPLETED FOR L29
```

PROCESSING COMPLETED FOR L40
L41 8 DUP REM L29 L40 (4 DUPLICATES REMOVED)

=> d 1-8 an ti au so ab; fil home

L41 ANSWER 1 OF 8 COPYRIGHT 1993 BIOSIS
AN 89:438231 BIOSIS
TI PACKAGING AND TRANSFER OF A MARKER GENE BY HIV VECTOR PARTICLES.
AU CLAVEL F; DANOS O; ALIZON M
SO MORISSET, R. A. (ED.). VE CONFERENCE INTERNATIONALE SUR LE SIDA: LE DEFI SCIENTIFIQUE ET SOCIAL; V INTERNATIONAL CONFERENCE ON AIDS: THE SCIENTIFIC AND SOCIAL CHALLENGE; MONTREAL, QUEBEC, CANADA, JUNE 4-9, 1989. 1262P. INTERNATIONAL DEVELOPMENT RESEARCH CENTRE: OTTAWA, ONTARIO, CANADA. ILLUS. PAPER. 0 (0). 1989. 583. ISBN: 0-662-56670-X

L41 ANSWER 2 OF 8 COPYRIGHT 1993 NLM
AN 87287230 MEDLINE
TI Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses.
AU Chakrabarti L; Guyader M; Alizon M; Daniel MD; Desrosiers RC; Tiollais P; Sonigo P
SO Nature, (1987 Aug 6-12) 328 (6130) 543-7
Journal code: NSC ISSN: 0028-0836
AB Because of the growing incidence of AIDS (acquired immune deficiency syndrome), the need for studies on animal models is urgent. Infection of chimpanzees with the retroviral agent of human AIDS, the human immunodeficiency virus (HIV), will have only limited usefulness because chimpanzees are in short supply and do not develop the disease. Among non-human primates, both type D retroviruses and lentiviruses can be responsible for immune deficiencies. The D-type retroviruses, although important pathogens in macaque monkey colonies, are not satisfactory as a model because they differ in genetic structure and pathophysiological properties from the human AIDS viruses. The simian lentivirus, previously referred to as simian T-cell lymphotropic virus type III (STLV-III), now termed simian immunodeficiency virus (SIV) is related to HIV by the antigenicity of its proteins and in its main biological properties, such as cytopathic effect and tropism for CD4-bearing cells. Most importantly, SIV induces a disease with remarkable similarity to human AIDS in the common rhesus macaques, which therefore constitute the best animal model currently available. Natural or experimental infection of other monkeys such as African green monkeys or sooty mangabeys has not yet been associated with disease. Molecular approaches of the SIV system will be needed for biological studies and development of vaccines that could be tested in animals. We have cloned and sequenced the complete genome of SIV isolated from a naturally infected macaque that died of AIDS. This SIVMAC appears genetically close to the agent of AIDS in West Africa, HIV-2, but the divergence of the sequences of SIV and HIV-2 is greater than that previously observed between HIV-1 isolates.

L41 ANSWER 3 OF 8 COPYRIGHT 1993 NLM
AN 87090385 MEDLINE
TI Molecular cloning and polymorphism of the human immune deficiency virus type 2.
AU Clavel F; Guyader M; Guetard D; Salle M; Montagnier L; Alizon M
SO Nature, (1986 Dec 18-31) 324 (6098) 691-5
Journal code: NSC ISSN: 0028-0836

AB We recently reported the isolation of a novel retrovirus, the human immune deficiency virus type 2 (HIV -2, previously named LAV-2), from patients with acquired immune deficiency syndrome (AIDS) originating from West Africa. This virus is related to HIV-1, the causative agent of the AIDS epidemic now spreading in Central and East Africa, as well as the USA and Europe (see ref. 3 for review) both by its morphology and by its tropism and in vitro cytopathic effect on CD4 (T4) positive cell lines and lymphocytes. But preliminary hybridization experiments indicated that there are substantiated differences between the sequences of the two genomes. Furthermore, the proteins of HIV-1 and HIV-2 have different sizes and their serological cross-reactivity is restricted to the major core protein, as the envelope glycoproteins of HIV-2 are not immunoprecipitated by HIV-1-positive sera. We now report the molecular cloning of the complete 9.5-kilobase (kb) genome of HIV-2, the observation of restriction site polymorphism between different isolates, and a preliminary analysis of the relationship of HIV-2 with other human and simian retroviruses.

L41 ANSWER 4 OF 8 COPYRIGHT 1993 BIOSIS DUPLICATE 1
AN 86:379265 BIOSIS
TI LYMPHADENOPATHY-ACQUIRED IMMUNE DEFICIENCY SYNDROME VIRUS GENETIC ORGANIZATION AND RELATIONSHIP TO ANIMAL LENTIVIRUSES.
AU ALIZON M; MONTAGNIER L
SO ANTICANCER RES 6 (3 PART B). 1986. 403-412. CODEN: ANTRD4 ISSN: 0250-7005
AB This article presents data obtained by our group in the molecular characterization of the probable agent of the acquired immune deficiency syndrome (AIDS), the lymphadenopathy/AIDS virus (LAV). Molecular cloning and complete nucleotide sequencing of LAV allows a detailed comparison with other AIDS virus isolates, as well as other human and animal retroviruses. We have now molecular evidence that the AIDS virus is closely related to visna virus, prototype of the lentiviruses, whereas the other human retroviruses, i.e., human T-cell leukemia viruses type I and II (HTLV-I and II), are quite remote in the evolution.

L41 ANSWER 5 OF 8 COPYRIGHT 1993 BIOSIS DUPLICATE 2
AN 86:377557 BIOSIS
TI GENETIC VARIABILITY OF THE ACQUIRED IMMUNE DEFICIENCY SYNDROME VIRUS NUCLEOTIDE SEQUENCE ANALYSIS OF TWO ISOLATES FROM AFRICAN PATIENTS.
AU ALIZON M; WAIN-HOBSON S; MONTAGNIER L; SONIGO P
SO CELL 46 (1). 1986. 63-74. CODEN: CELLB5 ISSN: 0092-8674
AB To define further the genetic variability of the human AIDS retrovirus, we have cloned and sequenced the complete genomes of two isolates obtained from Zairian patients. Their genetic organization is identical with that of isolates from Europe and North America, confirming a common evolutionary origin. However, the comparison of homologous proteins from these different isolates reveals a much greater extent of genetic polymorphism than previously observed. It is nevertheless possible to define conserved domains in the viral proteins, especially in the envelope, that could be of interest for the understanding of the molecular mechanisms of viral pathogenicity and for the development of diagnostic and therapeutic reagents.

L41 ANSWER 6 OF 8 COPYRIGHT 1993 BIOSIS DUPLICATE 3
AN 86:99324 BIOSIS

TI NUCLEOTIDE SEQUENCE OF THE VISNA LENTIVIRUS RELATIONSHIP TO THE AIDS ACQUIRED IMMUNE DEFICIENCY SYNDROME VIRUS.
AU SONIGO P; ALIZON M; STASKUS K; KLATZMANN D; COLE S; DANOS O; RETZEL E; TIOLLAIS P; HAASE A; WAIN-HOBSON S
SO CELL 42 (1). 1985. 369-382. CODEN: CELLB5 ISSN: 0092-8674
AB We have determined the complete 9202 nucleotide sequence of the visna lentivirus. The deduced genetic organization most closely resembles that of the AIDS retrovirus in that there is a novel central region separating pol and env. Moreover, there is a close phylogenetic relationship between the conserved reverse transcriptase and endonuclease/integrase domains of the visna and AIDS viruses. These findings support the inclusion of the AIDS virus in the retroviral subfamily Lentivirinae.

L41 ANSWER 7 OF 8 COPYRIGHT 1993 BIOSIS DUPLICATE 4
AN 85:296617 BIOSIS
TI NUCLEOTIDE SEQUENCE OF THE ACQUIRED IMMUNE DEFICIENCY SYNDROME VIRUS LYMPHADENOPATHY-ASSOCIATED VIRUS.
AU WAIN-HOBSON S; SONIGO P; DANOS O; COLE S; ALIZON M
SO CELL 40 (1). 1985. 9-18. CODEN: CELLB5 ISSN: 0092-8674
AB The complete 9193-nucleotide sequence of the probable causative agent of AIDS [acquired immune deficiency syndrome], lymphadenopathy-associated virus (LAV), was determined. The deduced genetic structure is unique: it shows, in addition to the retroviral gag, pol and env genes, 2 novel open reading frames termed Q and F. Remarkably, Q is located between pol and env and F is half-encoded by the U3 element of the LTR [long terminal repeat]. The data place LAV apart from the previously characterized family of human T cell leukemia/lymphoma viruses.

L41 ANSWER 8 OF 8 COPYRIGHT 1993 NLM
AN 85086249 MEDLINE
TI Molecular cloning of lymphadenopathy-associated virus.
AU Alizon M; Sonigo P; Barre-Sinoussi F; Chermann JC; Tiollais P; Montagnier L; Wain-Hobson S
SO Nature, (1984 Dec 20-1985 Jan 2) 312 (5996) 757-60
Journal code: NSC ISSN: 0028-0836
AB Lymphadenopathy-associated virus (LAV) is a human retrovirus first isolated from a homosexual patient with lymphadenopathy syndrome, frequently a prodrome or a benign form of acquired immune deficiency syndrome (AIDS). Other LAV isolates have subsequently been recovered from patients with AIDS or pre-AIDS and all available data are consistent with the virus being the causative agent of AIDS. The virus is propagated on activated T lymphocytes and has a tropism for the T-cell subset OKT4 (ref. 6), in which it induces a cytopathic effect. The major core protein of LAV is antigenically unrelated to other known retroviral antigens. LAV-like viruses have more recently been independently isolated from patients with AIDS and pre-AIDS. These viruses, called human T-cell leukaemia/lymphoma virus type III (HTLV-III) and AIDS-associated retrovirus (ARV), seem to have many characteristics in common with LAV and probably represent independent isolates of the LAV prototype. We have sought to characterize LAV by the molecular cloning of its genome. A cloned LAV complementary DNA was used to screen a library of recombinant phages constructed from the genomic DNA of LAV-infected T lymphocytes. Two families of clones were characterized which differ in a restriction site. The viral genome is longer than any other human retroviral genome (9.1-9.2 kilobases).

with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Schüpbach, J., Popovic, M., Gilden, R. V., Gonda, M. A., Sarngadharan, M. G., and Gallo, R. C. (1984). Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. Science *224*, 503–505.

Schwartz, D. E., Tizard, R., and Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. Cell *32*, 853–869.

Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA *80*, 3618–3622.

Shaw, G. M., Hahn, B. H., Arya, S. K., Groopman, J. E., Gallo, R. C., and Wong-Staal, F. (1984). Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. Science *226*, 1165–1171.

Shimotohno, K., and Temin, H. M. (1982). Spontaneous variation and synthesis in the U3 region of the long terminal repeat of an avian retrovirus. J. Virol. *41*, 163–171.

Shimotohno, K., Golde, D. M., Miwa, M., Sugimura, T., and Chen, I. S. Y. (1984). Nucleotide sequence analysis of the long terminal repeat of human T-cell leukemia virus type II. Proc. Natl. Acad. Sci. USA *81*, 1079–1083.

Shinnick, T. M., Lerner, R. A., and Sutcliffe, J. G. (1981). Nucleotide sequence of Moloney murine leukemia virus. Nature *293*, 543–548.

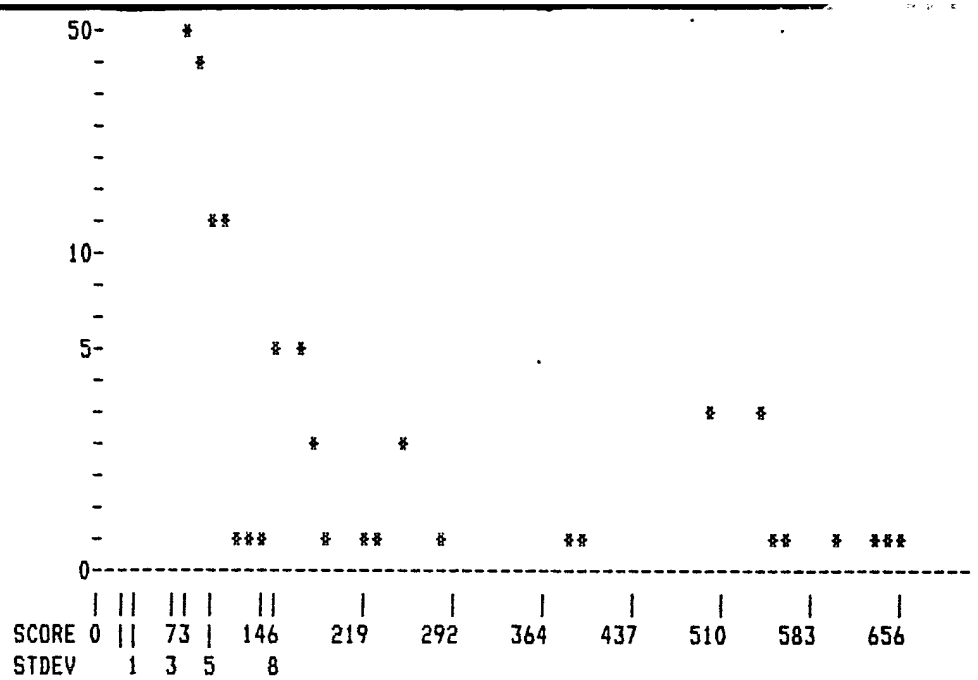
Srinivasan, A., Reddy, E. P., Dunn, C. Y., and Aaronson, S. A. (1984). Molecular dissection of transcriptional control elements with the long terminal repeat of retrovirus. Science *223*, 288–289.

Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucl. Acids. Res. *10*, 4731–4751.

Temin, H. (1981). Structure, variation and synthesis of retrovirus long terminal repeat. Cell *27*, 1–3.

Weinberg, R. A. (1982). Fewer and fewer oncogenes. Cell *30*, 3–9.

```
> 0 <
0| |0  IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file railey-000-716-ngs.res made by shears on Mon 26 Apr 93 14:38:23-PDT.


Query sequence being compared:RAILEY-000-716.SEQ (1-696)
Number of sequences searched:                20342
Number of scores above cutoff:                4112

      Results of the initial comparison of RAILEY-000-716.SEQ (1-696) with:
   Data bank : N-GeneSeq 9, all entries

 10000-
      *
N     -
U 5000-*
M     - *
B     -  *
E     -
R     -
      -   *
0     -
F 1000-
      -
S     -    *
E  500-
Q     -
U     -
E     -
N     -     *
C     -
E     -
S  100-
      -
      -
```



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	10
Optimized scores to save	0	Display context	50

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	21	17	15.71

Times:	CPU	Total Elapsed
	00:04:07.98	00:08:22.00

Number of residues: 12982290
Number of sequences searched: 20342
Number of scores above cutoff: 4112

Cut-off raised to 10.
Cut-off raised to 17.
Cut-off raised to 25.
Cut-off raised to 30.
Cut-off raised to 33.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 40 standard deviations above mean ****					
1. 014751	HIV-1(MN) env protein-encodin	9739	656	663	40.42	0

	**** 39 standard deviations above mean ****					
2. Q22488	HIV-1 proviral clone pNL4-3.	9709	640	640	39.40	0
	**** 38 standard deviations above mean ****					
3. N60240	HTLV-III virus (HIV virus) DN	9745	633	623	38.96	0
	**** 36 standard deviations above mean ****					
4. Q14752	HIV-1(MN-ST1) env protein-enc	9746	602	641	36.98	0
	**** 33 standard deviations above mean ****					
5. N60365	Sequence of LAV virus genome	9193	554	554	33.93	0
6. N60288	Sequence of the HTLV-III geno	9213	547	547	33.48	0
7. N60476	Sequence of lymphadenopathy-a	9088	542	542	33.17	0
8. Q15226	HIV-1 TAT mRNA.	1833	541	545	33.10	0
9. N71016	Sequence of LAV/HTLV III enve	4020	541	541	33.10	0
	**** 30 standard deviations above mean ****					
10. N80436	Entire sequence of LAV EL I	9236	502	502	30.62	0
11. Q06635	Complete sequence of HIV 1-ND	9143	499	499	30.43	0
12. N60140	Sequence of ARV-2 (9B) cDNA i	9737	493	652	30.05	0
	**** 23 standard deviations above mean ****					
13. Q11943	Nucleotide sequence of HIV-1	9192	394	513	23.74	0
	**** 22 standard deviations above mean ****					
14. N80437	Entire sequence of LAV MA L	9229	382	470	22.98	0
	**** 16 standard deviations above mean ****					
15. Q14753	HIV-1 BA-L clone.	3807	282	282	16.61	0
	**** 14 standard deviations above mean ****					
16. N80890	Sequence of cDNA clone HIV-2	9633	246	342	14.32	0
17. N92119	Sequence of clone HIV-2 SBL/I	9693	246	342	14.32	0
	**** 13 standard deviations above mean ****					
18. N71017	Sequence of LAV/HTLV III gag	5340	235	235	13.62	0
	**** 12 standard deviations above mean ****					
19. N90824	HIV LTR gene structure.	718	221	228	12.73	0
	**** 10 standard deviations above mean ****					
20. Q21163	COS cell expression vector pi	2932	192	377	10.89	0
	**** 9 standard deviations above mean ****					
21. Q02829	DNA complementary to simian i	9170	177	313	9.93	0
22. Q20616	ROD HIV-2 isolate complete ge	9672	175	351	9.80	0
23. N92768	HIV-2 variant HIV-D194 clone.	9473	174	290	9.74	0
24. N80859	Sequence of entire HIV-2 ROD	9643	173	349	9.68	0
25. N91774	Entire HIV-2/ST provirus DNA	9822	168	345	9.36	0
26. N92618	Portion of the HIV-3 retrovir	360	167	220	9.29	0
	**** 8 standard deviations above mean ****					
27. Q24802	SIVmac239 nef-deletion.	10097	151	343	8.28	0
28. Q22487	SIVmac239 proviral genome.	10279	151	343	8.28	0
29. N90375	DNA sequence of expression ve	1143	149	375	8.15	0
30. N90606	piH3M vector (ATCC 67,633) DN	3353	148	381	8.08	0
31. Q21166	Expression vector piH3M.	3900	148	381	8.08	0
	**** 7 standard deviations above mean ****					
32. N92769	HIV-2 variant HIV-D205 clone	324	137	175	7.38	0
	**** 6 standard deviations above mean ****					
33. Q10203	Sequence of simian immunodeff	9215	128	331	6.81	0
34. Q02830	cDNA to HIV-2 RNA.	9360	117	275	6.11	0
	**** 5 standard deviations above mean ****					
35. Q13189	Synthetic TAR sequence.	120	109	114	5.60	0
36. N93063	Sequence encoding hybrid prot	1383	105	266	5.35	0

37

37	N50333	Sequence of exons I and II an	986	104	292	5.28	0
38.	Q20532	Sequence of clone lambdaAPCP1	2256	103	303	5.22	0
39.	Q10014	Clone lambda APCP168i4 of bet	2256	103	303	5.22	0
40.	N80604	Lambda APCP168i4,amino acids	2256	103	303	5.22	0

1. RAILEY-000-716.SEQ (1-696)
   Q14751 HIV-1(MN) env protein-encoding sequence.

ID Q14751 standard; DNA; 9739 BP.
AC Q14751;
DT 05-FEB-1992 (first entry)
DE HIV-1(MN) env protein-encoding sequence.

KW human immunodeficiency virus; United States; MN isolate; AIDS;
KW envelope protein; ss.
OS Human immunodeficiency virus-1 (MN).
FH Key Location/Qualifiers
FT CDS 6240..8810
FT /*tag= a
FT /product= env
PN US7599491-A.
PD 15-OCT-1991.
PF 17-OCT-1990; 183830.
PR 17-OCT-1990; US-599491.
PA (USSH) NAT INST OF HEALTH.
PI Reitz M;
DR WPI; 91-346752/47.
DR P-PSDB; R14903.
PT US HIV-1 isolates MN-ST1 and BA-L, ENV protein and DNA - are
PT useful in therapeutics, vaccines and diagnostic tests
PS Example 1; Fig 2; 61pp; English.
CC The permuted circular unintegrated viral DNA representing the
CC complete HIV-1(MN) genome was cloned into the EcoRI site of lambda
CC gtWES.lambda B DNA from total DNA of H9 cells producing HIV-1 (MN).
CC This clone was designated lambda MN-PH1; it was subcloned in M13mp18
CC and M13mp19 and the DNA sequence of the entire clone was obtained.
CC The four "OTHERS" in the sequence represent bases which are
CC illegible in the specification. The amino acid sequence of the env
CC protein was deduced from this sequence and the env gene was
CC subcloned so that recombinant production of the env protein was
CC possible.
SQ Sequence 9739 BP; 3457 A; 1774 C; 2313 G; 2191 T;
SQ 4 Others;

Initial Score = 656 Optimized Score = 663 Significance = 40.42
Residue Identity = 96% Matches = 665 Mismatches = 21
Gaps = 3 Conservative Substitutions = 0

```
          10        20        30        40        50        60        70
   GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
         X       10        20        30        40        50        60

       80        90       100       110       120       130       140
   GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
   ||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||||
   GGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
      70        80        90       100       110       120       130

      150       160       170       180       190       200       210
   TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
   |||||||||||||||||||||||||| ||| |||||||| ||||| ||||||||||||||||||||||||||
   TACAAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAAGCCAACAAAGGAGAGAACACCAGCTTGTTACAC
   140       150       160       170       180       190       200

     220       230       240       250       260       270       280
   CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
   |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
   CCTGTGAGCCTGCATGGAATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
 210       220       230       240       250       260       270       280

   290       300       310       320       330       340       350       360
   TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
   ||||||||| |||||||||||||||||||||||||||||||||| |||||||||||||||||| |||
   TTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTCACATCGAGCTTGCTACAATGGA
         290       300       310       320       330       340       350

          370       380       390       400       410       420       430
```

```
   CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
   |||||||||||| |||||||||| |||||||||||||||| |||||||||||||||||||||||||| ||||
   CTTTCCGCTGGGGACTTTCCAGGTAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGC
       360        370        380        390        400        410        420

        440        450        460        470        480        490        500
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
   |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
     430        440        450        460        470        480        490

      510        520        530        540        550        560        570
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   500        510        520        530        540        550        560

    580        590        600        610        620        630        640
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
   ||| |||||||||||||| |||||||||||||||| |||||||| ||||||||||||||||||||||||||||
   GTTATGTGACTCTGGTAGCTAGAGATCCCTCAGATCCTTTTAGGCAGTGTGGAAAATCTCTAGCAGTGGCGC
 570        580        590        600        610        620        630        640

  650        660        670        680        690     X
   CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
   |||||||||||||||||||||||||  |||||||   ||||||||||||
   CCGAACAGGGACTTGAAAGCGAAAGAAAAACCA---GAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
         650        660        670          680     X 690        700        710

   CGCACGGCAAGAGGCGAGGGGCGGCG
          720        730
```

2. RAILEY-000-716.SEQ (1-696)
   Q22488 HIV-1 proviral clone pNL4-3.

```
 ID   Q22488 standard; DNA; 9709 BP.
 AC   Q22488;
 DT   06-JUL-1992  (first entry)
 DE   HIV-1 proviral clone pNL4-3.
 KW   AIDS; Acquired Immune Deficiency Syndrome; polymerase chain reaction;
 KW   PCR; site-directed mutagenesis; retrovirus; null-mutation; human; ss.
 OS   Human immunodeficiency virus.
 FH   Key             Location/Qualifiers
 FT   repeat_region   1..634
 FT   /*tag=  a
 FT   /rpt_type= TERMINAL
 FT   /note= "5'LTR"
 FT   repeat_unit     456..548
 FT   /*tag=  b
 FT   /standard_name= R
 FT   GC_signal       375..385
 FT   /*tag=  c
 FT   /standard_name= Sp1_binding_site
 FT   GC_signal       389..395
 FT   /*tag=  d
 FT   /standard_name= Sp1_binding_site
 FT   GC_signal       399..407
 FT   /*tag=  e
 FT   /standard_name= Sp1_binding_site
 FT   primer_bind     636..656
 FT   /*tag=  f
 FT   /standard_name= Lys_tRNA_pbs
 FT   CDS             790..2292
 FT   /*tag=  g
 FT   /product= gag
```

FT CDS 2087..5096
FT /*tag= h
FT /product= pol
FT /note= "NH2-terminal uncertain"
FT CDS 5041..5619
FT /*tag= i
FT /product= vif
FT CDS 5559..5849
FT /*tag= j
FT /product= vpr
FT CDS 6061..6306
FT /*tag= k
FT /product= vpu
FT exon 5830..6044
FT /*tag= l
FT /product= tat
FT /note= "full-length tat obtained by splicing"
FT exon 5969..6044
FT /*tag= m
FT /product= rev
FT /note= "full-length rev obtained by splicing"
FT exon 8369..8414
FT /*tag= n
FT /product= tat
FT /note= "see above"
FT exon 8369..8643
FT /*tag= o
FT /product= rev
FT /note= "see above"
FT CDS 6221..8785
FT /*tag= p
FT /product= env
FT CDS 8787..9407
FT /*tag= q
FT /product= nef
FT repeat_region 9076..9709
FT /*tag= r
FT /rpt_type= TERMINAL
FT /note= "3'LTR"
FT repeat_unit 9531..9624
FT /*tag= s
FT /standard_name= R
FT polyA_signal 9602..9607
FT /*tag= t
PN WO9200987-A.
PD 23-JAN-1992.
PF 10-JUL-1991; U04884.
PR 12-JUL-1990; US-551945.
PA (HARD) HARVARD COLLEGE.
PI Desrosiers RC.
DR WPI; 92-056816/07.
PT Primate lentivirus vaccine protecting against AIDS - and primate
PT lentiviruses and their DNA clones contg. null mutations, useful for
PT producing vaccine
PS Disclosure; Fig 3; 51pp; English.
CC The proviral clone pNL4-3 was used as the basis for creating the
CC null-mutations of the invention. The clone was described in
CC Adachi et al., J.Virol. 59:284, 1986. See G21079-G21086 for
CC examples of mutagenic primers for site-directed deletion of regions
CC of NL4-3.
SQ Sequence 9709 BP; 3421 A; 1759 C; 2365 G; 2161 T;
SQ 3 Others;

Initial Score = 640 Optimized Score = 640 Significance = 39.40
Residue Identity = 92% Matches = 640 Mismatches = 49
Gaps = 0 Conservative Substitutions = 0

```
         10        20        30        40        50        60        70
  GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
        ||||||||||||||   |||||| |||||||| ||||||||||||||||   ||||||||||||
        TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGNNNCACCACACACAA
        X       10        20        30        40        50        60

       80        90       100       110       120       130       140
  GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
  ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
  GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
     70        80        90       100       110       120       130

    150       160       170       180       190       200       210
  TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
  | |||| |||||||||||||| |||||  | |||||||||||||| |||||||||| | ||||||||||||
  TTCAAGTTAGTACCAGTTGAACCAGAGCAAGTAGAAGAGGCCAAATAAGGAGAGAAGAACAGCTTGTTACAC
  140       150       160       170       180       190       200

   220       230       240       250       260       270       280
  CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
  ||| |||||| |||||| ||||| ||||| ||| |||||||| |||| ||||| |||||||||||| |||||||
  CCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGCA
210       220       230       240       250       260       270       280

 290       300       310       320       330       340       350       360
  TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
  |||| |||| |||||||| |||||||||||||||||| |||  |||||||||||||||||| ||||||||||
  TTTCGTCACATGGCCCGACAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGA
       290       300       310       320       330       340       350

        370       380       390       400       410       420       430
  CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
  |||||||||||| ||||||||||||| |||||||||||||| |||||||||||||||||||||||||||| |
  CTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTAC
      360       370       380       390       400       410       420

      440       450       460       470       480       490       500
  ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
  ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
    430       440       450       460       470       480       490

     510       520       530       540       550       560       570
  CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
  CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTGTGCCCGTCT
  500       510       520       530       540       550       560

   580       590       600       610       620       630       640
  GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
570       580       590       600       610       620       630       640

 650       660       670       680       690     X
  CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
  |||||||||||||||||||||||||  || ||||||||| ||||||||
  CCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
       650       660       670       680       690       700       710

  CGCACGGCAAGAGGCGAGGGGCGGCG
      720       730
```

3. RAILEY-000-716.SEQ (1-696)

N60240 HTLV-III virus (HIV virus) DNA.

ID N60240 standard; DNA; 9745 BP.
AC N60240;
DT 01-JAN-1980 (first entry)
DE HTLV-III virus (HIV virus) DNA.
KW HTLV-III; HIV virus; AIDS; active immunization;
KW passive immunization; vaccine; ss.
OS HIV virus (HTLV-III).
FH Key Location/Qualifiers
FT CDS 786..2318
FT /*tag= a
FT /note= "gag protein open reading frame"
FT CDS 2078..5122
FT /*tag= b
FT /note= "pol protein open reading frame"
FT CDS 5037..5646
FT /*tag= c
FT /note= "sor protein open reading frame"
FT CDS 6230..8818
FT /*tag= d
FT /note= "env-lor protein open reading frame"
PN EP-185444-A.
PD 25-JUN-1986.
PF 10-OCT-1985; 307260.
PR 10-OCT-1984; US-659339.
PR 23-JAN-1985; US-693866.
PA (CENT-) CENTOCOR INC.
PI Chang NT;
DR WPI; 86-163443/26.
DR P-PSDB; P60346-49.
PT New immunoreactive HTLV-III polypeptide expressed by transformed
PT cells - and derived antibodies, useful for diagnosis of AIDS and
PT in active or passive immunisation
PS Disclosure; Fig. 3; 60pp; English.
CC HIV virus cDNA is cleaved with restriction endonucleases to produce
CC fragments coding for the specified proteins. The resulting proteins,
CC gag, pol, sor and env-lor, and antibodies against them are useful
CC for immunoassay of HIV virus, e.g. by sandwich type RIA. The
CC proteins may also be used in vaccines for active immunization.
SQ Sequence 9745 BP; 3434 A; 1782 C; 2363 G; 2166 T;

Initial Score = 633 Optimized Score = 623 Significance = 38.96
Residue Identity = 97% Matches = 626 Mismatches = 13
Gaps = 4 Conservative Substitutions = 0

```
                                                    X       10        20
                                                    GGGGGACTGGAAGGGCTAATTC
                                                    ||||||||||||||||||||||
   CAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTC
     9060      9070      9080      9090      9100    X 9110      9120

          30        40        50        60        70        80        90
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGA
   9130      9140      9150      9160      9170      9180      9190

      100       110       120       130       140       150       160
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
   |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
 9200      9210      9220      9230      9240      9250      9260      9270

    170       180       190       200       210       220       230
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGA--ACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAAT
```

```
CAGAGAAGTTAGAAGAAGCCAACAAAGGAGAGAACACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAAT
      9280      9290      9300      9310      9320      9330      9340

 240       250       260       270       280       290       300
GGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGA
GGATGACCC--GGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACATGGCCCGAGA
    9350      9360      9370      9380      9390      9400      9410

310       320       330       340       350       360       370       380
GCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCC
GCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCC
    9420      9430      9440      9450      9460      9470      9480

      390       400       410       420       430       440       450
AGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTG
AGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGCATATAAGCAGCTGCTTTTTG
  9490      9500      9510      9520      9530      9540      9550

     460       470       480       490       500       510       520
CCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGC
CCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAGCTAGGGAACCCACTGC
9560      9570      9580      9590      9600      9610      9620

   530       540       550       560       570       580       590
TTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACT
TTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACT
9630      9640      9650      9660      9670      9680      9690      9700

 600       610       620       630       640       650       660
AGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGC
AGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA
      9710      9720      9730      9740    X

670       680       690
GAAAGGGAAACCAGAGGAGCTCT
```

Defined by Cheng — which has BH10 supplemented with HXB2 sequences to complete a 5'LTR see cheng et al

4. RAILEY-000-716.SEQ (1-696)
   Q14752 HIV-1(MN-ST1) env protein-encoding sequence.

```
ID   Q14752 standard; DNA; 9746 BP.
AC   Q14752;
DT   05-FEB-1992  (first entry)
DE   HIV-1(MN-ST1) env protein-encoding sequence.
KW   human immunodeficiency virus; United States; MN isolate; AIDS;
KW   envelope protein; ss.
OS   Human immunodeficiency virus-1 (MN).
FH   Key             Location/Qualifiers
FT   CDS             6243..8806
FT   /*tag= a
FT   /product= env
PN   US7599491-A.
PD   15-OCT-1991.
PF   17-OCT-1990; 183830.
PR   17-OCT-1990; US-599491.
PA   (USSH ) NAT INST OF HEALTH.
PI   Reitz H;
DR   WPI; 91-346752/47.
DR   P-PSDB; R14904.
```

PT US HIV-1 isolates MN-ST1 and BA-L, ENV protein and DNA - are
PT useful in therapeutics, vaccines and diagnostic tests
PS Example 2; Fig 6; 61pp; English.
CC The infectious molecular clone, lambda MN-ST1, was obtained by
CC cloning integrated provirus from DNA purified from peripheral blood
CC lymphocytes infected with HIV-1(MN) and maintained in culture for
CC one month. The integrated proviral DNA was partially digested with
CC Sau3A to give fragments of 15-20 kb. The fragments were cloned in
CC EMBL3 and the entire sequence of the clone was determined.
SQ Sequence 9746 BP; 3465 A; 1752 C; 2355 G; 2174 T;

Initial Score = 602 Optimized Score = 641 Significance = 36.98
Residue Identity = 93% Matches = 645 Mismatches = 41
Gaps = 5 Conservative Substitutions = 0

```
          10        20        30        40        50        60        70
    GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
           |||| ||| ||||| |||||||| | |||||||| ||||||||||||||| ||||||||||||||
           TGGATGGGTTAATTTACTCCCAAAG-AGACAAGACATCCTTGATCTGTGGGTCTACCACACACAA
           X       10        20        30        40        50        60

         80        90       100       110       120       130       140
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
    ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
        70        80        90       100       110       120       130

       150       160       170       180       190       200       210
    TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
    | |||||||||||||||||||||||| ||| |||||||||||||||||||||||||| | ||||||||||
    TTCAAGCTAGTACCAGTTGAGCCAGAGAAGATAGAAGAGGCCAATAAAGGAGAGAACAACTGCTTGTTACAC
     140       150       160       170       180       190       200

     220       230       240       250       260       270       280
    CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
    ||| |||||| |||||| |||||||||| |||||||||||||||| ||||| || |||||||| |||||||
    CCTATGAGCCAGCATGGGATGGATGACCCGGAGAGAGAAGTGTTAGTGTGGAAGTCTGACAGCCACCTAGCA
   210       220       230       240       250       260       270       280

   290       300       310       320       330       340       350       360
    TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
    ||||| ||    |||||||||||||||||||||||| |||||||||||||||||||||  ||||||||||
    TTTCAGCATTATGCCCGAGAGCTGCATCCGGAGTACTACAAGAACTGCTGACATCGAGCTATCTACAAGGGA
           290       300       310       320       330       340       350

           370       380       390       400       410       420       430
    CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
    |||||||||||| |||||||||||||| |||||||||||||| || ||||||||||||||||||||||||||
    CTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACCGGGGAGTGGCGAGCCCTCAGATGCTGC
         360       370       380       390       400       410       420

         440       450       460       470       480       490       500
    ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
    |||||||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||||||||||
    ATATAAGCAGCTGCTTTCTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
       430       440       450       460       470       480       490

       510       520       530       540       550       560       570
    CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
     500       510       520       530       540       550       560

     580       590       600       610       620       630       640
    GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
    ||| ||| ||||||||| |||||||||||||||| |||||||| || ||||||||||||||||||||||||
```

```
   GTTATGTCACTCTGGTAGCTAGAGATCCCTCAGATCCTTTTAGGCA--GTGGAAAATCTCTAGCAGTGGCGC
  570       580       590       600       610         620       630

  650       660         670       680       690     X
   CCGAACAGGGAC--TTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
   ||||||||||||   |||||||||||| ||||||||||||||||||||||
   CCGAACAGGGACCTCTGAAAGCGAAAGAGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAG
  640       650       660       670       680       690       700       710

   CGCGCACGGCAAGAGGCGAGGGGCGGCG
         720       730
```

5. RAILEY-000-716.SEQ (1-696)
   N60365 Sequence of LAV virus genome .

```
ID   N60365 standard; cDNA; 9193 BP.
AC   N60365;
DT   20-AUG-1991  (first entry)
DE   Sequence of LAV virus genome .
KW   AIDS vaccine; diagnosis; immunoassay; HIV; HTLV-III; ss.
OS   Lymphadenopathy virus.
FH   Key             Location/Qualifiers
FT   CDS             312..1838
FT   /*tag=  a
FT   /product= gag
FT   CDS             1631..4642
FT   /*tag=  b
FT   /product= pol
FT   CDS             4554..5165
FT   /*tag=  c
FT   /product= ORF Q
FT   CDS             5746..8352
FT   /*tag=  d
FT   /product= env
FT   CDS             8324..8974
FT   /*tag=  e
FT   /product= ORF F
PN   WO8602383-A.
PD   24-APR-1986.
PF   18-OCT-1985; E00548.
PR   18-OCT-1984; FR-016013.
PR   16-NOV-1984; GB-029099.
PR   21-JAN-1985; GB-001473.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (INSP ) INST PASTEUR.
PI   Montagnier L, Krust B, Chamaret S, Clavel F, Chermann J-C,
PI   Barre-Sinoussi F, Alizon M, Sonigo P, Stewart C, Danos O,
PI   Wain-Hobson S.
DR   WPI; 86-119166/18.
DR   P-PSDB; P60419, P60420, P60421, P60422, P60423.
PT   Purified glyco:protein and peptide(s) - are recognised by sera contg.
PT   antibodies against lymphadenopathy virus and useful in detecting
PT   AIDS antibodies or in vaccines
PS   Disclosure; Fig 4; 75pp; English.
CC   The inventors claim a polypeptide which is recognised by sera of
CC   human origin contg. antibodies against the virus of
CC   lymphadenopathies (LAV) or acquired immune deficiency syndrome
CC   (AIDS). Also claimed are various peptides corresp. to the AA
CC   sequences deducible from proteins encoded by LAV DNA, defined by
CC   specific residues (e.g. 12-32, 37-46, 49-79, 88-153) in accordance
CC   with a formula given in the specification.
SQ   Sequence   9193 BP;    3278 A;    1652 C;    2216 G;    2047 T;

Initial Score    =    554  Optimized Score  =    554  Significance = 33.93
Residue Identity =    99%  Matches          =    554  Mismatches   =      4
```

Gaps = 0 Conservative Substitutions = 0

```
                                                    X       10        20
                                                    GGGGGACTGGAAGGGCTAATTC
                                                    ||||||||||||||||||||||
   CAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTC
    8590      8600      8610      8620      8630      8640      8650

         30        40        50        60        70        80        90
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
  8660      8670      8680      8690      8700      8710      8720

     100       110       120       130       140       150       160
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
8730      8740      8750      8760      8770      8780      8790      8800

    170       180       190       200       210       220       230
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
        8810      8820      8830      8840      8850      8860      8870

  240       250       260       270       280       290       300       310
   ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
      8880      8890      8900      8910      8920      8930      8940

         320       330       340       350       360       370       380
   TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
   |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
   TGCATCCGCAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAG
    8950      8960      8970      8980      8990      9000      9010

        390       400       410       420       430       440       450
   GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
   |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
   GGAGGCGTGGCCTGGGGGGGACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
  9020      9030      9040      9050      9060      9070      9080

     460       470       480       490       500       510       520
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
9090      9100      9110      9120      9130      9140      9150      9160

    530       540       550       560       570       580       590
   AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAG
   ||||||||||||||||||||||||||||||||
   AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCA
        9170      9180      9190  X

  600
   AGATCCCTCA
```

6. RAILEY-000-716.SEQ (1-696)
   N60288 Sequence of the HTLV-III genome.

ID N60288 standard; DNA; 9213 BP.
AC N60288;
DT 08-JUN-1991 (first entry)

```
DE   Sequence of the HTLV-III genome.
KW   HIV; LAV; AIDS; diagnosis; vaccine; ss.
OS   HTLV-IIIB/H9 cells (ATCC CRL 8543).
FH   Key             Location/Qualifiers
FT   repeat_region   1..96
FT   /*tag= a
FT   misc_feature    97..183
FT   /*tag= b
FT   /label= unique region
FT   CDS             336..731
FT   /*tag= c
FT   /product= gag
FT   CDS             732..1772
FT   /*tag= d
FT   /product= p24gag
FT   CDS             1639..4677
FT   /*tag= e
FT   /product= pol
FT   CDS             4622..5200
FT   /*tag= f
FT   /product= p'
FT   CDS             5802..7335
FT   /*tag= g
FT   /product= env
FT   CDS             7336..8373
FT   /*tag= h
FT   /product= gp41env
FT   CDS             8375..8995
FT   /*tag= i
FT   /product= E'
FT   misc_feature    8662..9117
FT   /*tag= j
FT   /label= unique region
FT   repeat_region   9118..9213
FT   /*tag= k
FT   polyA_signal    9090..9095
FT   /*tag= l
FT   polyA_signal    9190..9195
FT   /*tag= m
PN   EP-187041-A.
PD   09-JUL-1986.
PF   23-DEC-1985; 309454.
PR   24-DEC-1984; US-685272.
PR   04-DEC-1985; US-805069.
PA   (GETH ) GENENTECH INC.
PI   Capon DJ, Lasky LA;
DR   WPI; 86-177602/28.
DR   P-PSDB; P60309, P61507, P61504, P61514, P61515.
PT   Acquired immune deficiency syndrome polypeptide(s) - obtd. by
PT   molecular cloning etc. and used for diagnosis and in vaccines
PT   against virus disease
PS   Example; fig 2; 125pp; English.
CC   A comparison of N60287 with the cDNA of the HTLV-III genome
CC   revealed one particular clone, designated p7.11 which contained a
CC   DNA sequence encoding this peptide (P60308) sequence.  This approx.
CC   2.2 kilobase covers the precursor gag region and encodes, 5' to 3',
CC   p-12, p-15, p-24 a second p-15 protein, and approx. 300 extra base
CC   pairs 3' to the gag region (see N60288).
SQ   Sequence   9213 BP;    3297 A;    1656 C;    2217 G;    2043 T;

Initial Score    =    547  Optimized Score  =    547  Significance = 33.48
Residue Identity =    98%  Matches          =    547  Mismatches   =    10
Gaps             =      0  Conservative Substitutions              =     0

                                                    X       10        20
                                                    GGGGGACTGGAAGGGCTAATTC
```

```
                                                          ||||||||||||||||||||||
   CAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTC
   8610      8620      8630      8640      8650      8660      8670

         30        40        50        60        70        80        90
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
 8680      8690      8700      8710      8720      8730      8740      8750

      100       110       120       130       140       150       160
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
   |||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTACACACCAGGACCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
         8760      8770      8780      8790      8800      8810      8820

    170       180       190       200       210       220       230
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
   |||||||||||||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||||
   CAGATAAGGTAGAAGAGGCCAACAAAGGAGAGAACACCAGCTTGTTACACCCTGTGACCCTGCATGGAATGG
       8830      8840      8850      8860      8870      8880      8890

  240       250       260       270       280       290       300       310
   ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
   ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
     8900      8910      8920      8930      8940      8950      8960

         320       330       340       350       360       370       380
   TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
   |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||||
   TGCATCCGGAGTACTTCAAGAACTGCTGATATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAG
   8970      8980      8990      9000      9010      9020      9030

       390       400       410       420       430       440       450
   GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
   |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
   GGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
 9040      9050      9060      9070      9080      9090      9100      9110

     460       470       480       490       500       510       520
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
   |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||
   TGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGAGAACCCACTGCTT
         9120      9130      9140      9150      9160      9170      9180

    530       540       550       560       570       580       590
   AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAG
   ||||||||||||||||||||||||||||||||
   AAGCCTCAATAAAGCTTGCCTTGAGTGCTTC
       9190      9200      9210  X

  600
   AGATCCCTC
```

7. RAILEY-000-716.SEQ (1-696)
   N60476 Sequence of lymphadenopathy-associated virus (LAV)

```
ID   N60476 standard; cDNA; 9088 BP.
AC   N60476;
DT   24-AUG-1991  (first entry)
DE   Sequence of lymphadenopathy-associated virus (LAV) genome in lambda-
DE   J19.
KW   HTLV-III; human T-cell leukemia/lymphoma virus type III; ARV; AIDS;
KW   associated retrovirus; HIV; ARC; probe; diagnosis; ss.
```

OS Lymphadenopathy-associated virus.
PN WO8601827-A.
PD 27-MAR-1986.
PF 19-SEP-1955; 007200.
PR 19-SEP-1984; GB-023659.
PA (INSP) INST PASTEUR.
PA (CNRS) CENT NAT RECH SCIENTIFIQU.
PI Alizon M, Barre Sinoussi F, Sonigo P, Tiollais P, Chermann JC,
PI Montagnier L, Wainhobson S;
DR WPI; 86-094080/14.
PT Cloned DNA contg. fragment hybridised with genomic RNA or LAV -
PT used for detection of lymphadenopathy-associated virus
PS Disclosure; Fig 4-11; 24pp; English.
CC THe inventors claim a DNA SQ which is hybridizable with the genomic
CC RNA of the LAV viruses. Specifically claimed are SQs which code for
CC the envelope proteins, polymerase and core proteins. Also claimed
CC is a probe for the in vitro detection of LAV. N60476 was prepd.
CC from virions from FR8, an immortalized permanent LAV producing B-
CC lymphocyte line.
SQ Sequence 9088 BP; 3257 A; 1624 C; 2185 G; 2022 T;

Initial Score = 542 Optimized Score = 542 Significance = 33.17
Residue Identity = 99% Matches = 542 Mismatches = 1
Gaps = 0 Conservative Substitutions = 0

```
                                                  X       10        20
                                                  GGGGGACTGGAAGGGCTAATTC
                                                  ||||||||||||||||||||||
    CAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTC
     8500      8510      8520      8530      8540      8550      8560

          30        40        50        60        70        80        90
    ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
   8570      8580      8590      8600      8610      8620      8630

       100       110       120       130       140       150       160
    ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
 8640      8650      8660      8670      8680      8690      8700      8710

     170       180       190       200       210       220       230
    CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
    |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
    CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGACCCTGCATGGAATGG
         8720      8730      8740      8750      8760      8770      8780

   240       250       260       270       280       290       300       310
    ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
       8790      8800      8810      8820      8830      8840      8850

          320       330       340       350       360       370       380
    TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
     8860      8870      8880      8890      8900      8910      8920

         390       400       410       420       430       440       450
    GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
   8930      8940      8950      8960      8970      8980      8990
```

```
      460       470       480       490       500       510       520
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
9000      9010      9020      9030      9040      9050      9060      9070

    530       540  X    550       560       570       580       590
   AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTA
   ||||||||||||||||||
   AAGCCTCAATAAAGCTT
        9080       X
```

8. RAILEY-000-716.SEQ (1-696)
   Q15226 HIV-1 TAT mRNA.

```
 ID   Q15226 standard; mRNA; 1833 BP.
 AC   Q15226;
 DT   11-MAR-1992  (first entry)
 DE   HIV-1 TAT mRNA.
 KW   Retrovirus; treatment; oligonucleotide; anti-sense; binding; ss.
 OS   Synthetic.
 PN   WO9118004-A.
 PD   28-NOV-1991.
 PF   22-APR-1991; U02734.
 PR   11-MAY-1990; US-521907.
 PA   (ISIS-) ISIS PHARM INC.
 PI   Ecker DJ;
 DR   WPI; 91-369176/50.
 PT   Anti-sense DNA capable of binding HIV virus TAT mRNA in human
 PT   cells - for treatment of retroviral disease e.g. AIDS
 PS   Disclosure; Fig 1; 24pp; English.
 CC   The oligonucleotides represented in Q15220-25 are capable of
 CC   binding at least a portion of tat mRNA of HIV.  They can be used to
 CC   treat HIV and other human retroviruses.  It is partic. effective
 CC   therapeutically because particular sites of the RNA of HIV or other
 CC   RNA are targeted e.g. the tat mRNA.
 SQ   Sequence   1833 BP;    525 A;    408 C;    510 G;    390 U;

Initial Score    =    541  Optimized Score  =    545  Significance = 33.10
Residue Identity =    73%  Matches          =    546  Mismatches   =    29
Gaps             =      1  Conservative Substitutions              =     0

                                                    X       10        20
                                                    GGGGGACTGGAAGGGCTAATTC
                                                    ||||||||||||||||||||||
   CAAUGACUUACAAGGCAGCUGUAGAUCUUAGCCACUUUUUAAAAGAAAAGGGGGGACUGGAAGGGCUAAUUC
  1210      1220      1230      1240      1250      1260      1270

         30        40        50        60        70        80        90
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACUCCCAACGAAGACAAGAUAUCCUUGAUCUGUGGAUCUACCACACACAAGGCUACUUCCCUGAUUAGCAGA
1280      1290      1300      1310      1320      1330      1340      1350

      100       110       120       130       140       150       160
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
   |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
   ACUACACACCAGGGCCAGGGAUCAGAUAUCCACUGACCUUUGGAUGGUGCUACAAGCUAGUACCAGUUGAGC
        1360      1370      1380      1390      1400      1410      1420

    170       180       190       200       210       220       230
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
   |||| ||| |||||||| |||||| ||||||||||||||||||||||||||||||||||| ||||||||||||
   CAGAGAAGUUAGAAGAAGCCAACAAAGGAGAGAACACCAGCUUGUUACACCCUGUGAGCCUGCAUGGAAUGG
```

```
      1430       1440       1450       1460       1470       1480       1490

  240        250        260        270        280        290        300        310
   ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
   ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
   AUGACCCGGAGAGAGAAGUGUUAGAGUGGAGGUUUGACAGCCGCCUAGCAUUUCAUCACAUGGCCCGAGAGC
    1500       1510       1520       1530       1540       1550       1560

          320        330        340        350        360        370        380
   TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
   UGCAUCCGGAGUACUUCAAGAACUGCUGACAUCGAGCUUGCUACAAGGGACUUUCCGCUGGGGACUUUCCAG
  1570       1580       1590       1600       1610       1620       1630

       390        400        410        420        430        440        450
   GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
   |||||||||||||||||| |||||||||||||||||||||| |||||||||||||||||||||||||||||
   GGAGGCGUGGCCUGGGCGGGACUGGGGAGUGGCGAGCCCUCAGAUCCUGCAUAUAAGCAGCUGCUUUUUGCC
 1640       1650       1660       1670       1680       1690       1700       1710

      460        470        480        490        500        510        520
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
   |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
   UGUACUGGGUCUCUCUGGUUAGACCAGAUCUGAGCCUGGGAGCUCUCUGGCUAACUAAGGAACCCACUGCUU
        1720       1730       1740       1750       1760       1770       1780

    530        540        550        560        570     X  580        590
   AAGCCTCAATAAAGCTTGCCTTGAGTGCT-TCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTA
   |||||||||||||||||||||||||||||| ||||  |
   AAGCCUCAAUAAAGCUUGCCUUGAGUGCUGUCAAAAAAAAAAAAAAAAAAA
      1790       1800       1810       1820       1830  X

   600        610        620
   GAGATCCCTCAGACCCTTTTAGTCAGTG
```

9. RAILEY-000-716.SEQ (1-696)
   N71016 Sequence of LAV/HTLV III envelope gene (env).

```
ID   N71016 standard; DNA; 4020 BP.
AC   N71016;
DT   23-APR-1991  (first entry)
DE   Sequence of LAV/HTLV III envelope gene (env).
KW   Glycoprotein gp 110; gp 41; AIDS vaccine; diagnosis; ss.
OS   LAV/HTLV III.
FH   Key             Location/Qualifiers
FT   CDS             487..3072
FT   /*tag=  a
FT   /note= "A recombinant virus contg. this SQ is
FT   claimed"
PN   WO8702038-A.
PD   09-APR-1987.
PF   24-SEP-1986; 022987.
PR   25-SEP-1985; US-779909.
PR   27-MAR-1986; US-842984.
PR   09-SEP-1986; US-905217.
PA   (ONCO-) ONCOGEN.
PA   (HUSS/) HU S L.
PI   Hu SL, Purchio AF, Madisen L;
DR   WPI; 87-108683/15.
DR   P-PSDB; P70665.
PT   New recombinant viruses for directing expression of peptide(s)
PT   etc. - useful in vaccines for protecting humans against AIDS
PT   caused by LAV/HTLV III
PS   Disclosure; Fig 2; 165pp; English.
CC   Recombinant Ac-NPV carrying the chimeric LAV/HTLV III env gene was
```

CC used to infect Sf9 cells in tissue culture. The proteins produced on
CC cultivation were immunoreactive with AIDS patient serum as well as
CC with monoclonal antibodies which define LAV/HTLV III envelope
CC glycoproteins gp. 110 and gp. 41. A recombinant DNA vector
CC comprising ps-env 1,2,5 or7 pv-gag1, pAc-gag1 or pAc-env 5, is
CC claimed.
SQ Sequence 4020 BP; 1352 A; 734 C; 990 G; 944 T;

Initial Score = 541 Optimized Score = 541 Significance = 33.10
Residue Identity = 99% Matches = 541 Mismatches = 4
Gaps = 0 Conservative Substitutions = 0

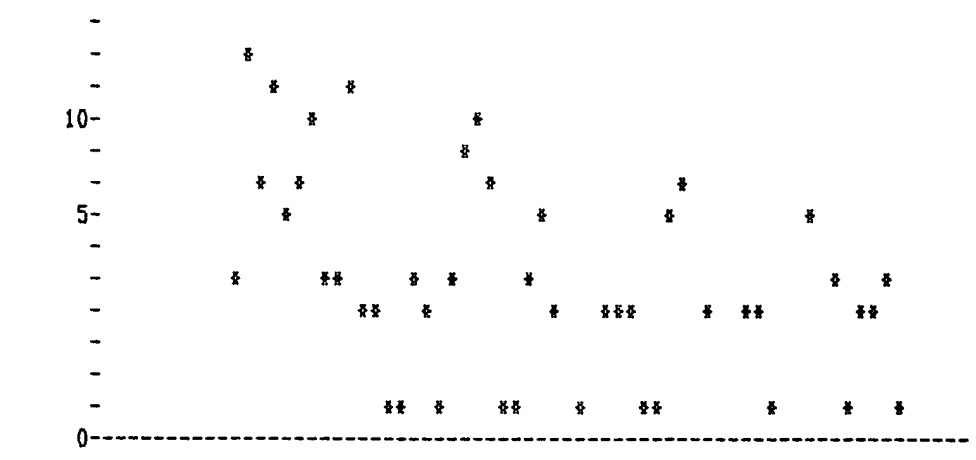
```
                                                X       10        20
                                                GGGGGACTGGAAGGGCTAATTC
                                                |||||||||||||||||||||||
   CAATGACTTACAAGGCAGCTCTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTC
    3430      3440      3450      3460      3470      3480      3490

         30        40        50        60        70        80        90
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
  3500      3510      3520      3530      3540      3550      3560

      100       110       120       130       140       150       160
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
3570      3580      3590      3600      3610      3620      3630      3640

    170       180       190       200       210       220       230
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
        3650      3660      3670      3680      3690      3700      3710

  240       250       260       270       280       290       300       310
   ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
   |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
   ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTACGATTTCATCACGTGGCCCGAGAGC
      3720      3730      3740      3750      3760      3770      3780

         320       330       340       350       360       370       380
   TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
   TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAG
    3790      3800      3810      3820      3830      3840      3850

        390       400       410       420       430       440       450
   GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
   |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
   GGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
  3860      3870      3880      3890      3900      3910      3920

     460       470       480       490       500       510       520
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
3930      3940      3950      3960      3970      3980      3990      4000

    530       540    X  550       560       570       580       590
   AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAAC
   ||||||||||||||||||||
   AAGCCTCAATAAAGCTTGC
        4010      4020
```

10. RAILEY-000-716.SEQ (1-696)
N80436 Entire sequence of LAV EL I

```
ID   N80436 standard; cDNA; 9236 BP.
AC   N80436;
DT   16-DEC-1990  (first entry)
DE   Entire sequence of LAV EL I
KW   HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation; ss.
OS   Lymphadenopathy associated virus EL I.
FH   Key             Location/Qualifiers
FT   misc_feature    1..98
FT   /*tag= a
FT   /label=R
FT   misc_feature    99..182
FT   /*tag= b
FT   /label=U5
FT   misc_feature    8683..9138
FT   /*tag= c
FT   /label=U3
FT   misc_feature    9139..9236
FT   /*tag= d
FT   /label=R
FT   CDS             336..1835
FT   /*tag= e
FT   /label=GAG, P80884
FT   CDS             1634..4699
FT   /*tag= f
FT   /label=POL, P81854
FT   CDS             4647..5222
FT   /*tag= g
FT   /label=Q, P81855
FT   CDS             5165..5452
FT   /*tag= h
FT   /label=R, P81856
FT   CDS             5436..5651
FT   /*tag= i
FT   /label=S, P81857
FT   CDS             5830..8388
FT   /*tag= j
FT   /label=ENV, P81858
FT   CDS             8393..9010
FT   /*tag= k
FT   /label=F, P81859
PN   WO8707906-A.
PD   30-DEC-1987.
PF   22-JUN-1987; E00326.
PR   23-JUN-1986; EP-401380.
PA   (INSP) Inst Pasteur.
PI   Alizon M, Sonigo P, Wain-Hobson S, Montagnier L;
DR   WPI; 88-014396/02.
DR   P-PSDB; P80884, P81854, P81855, P81856, P81857, P81858, P81859.
PT   New variants of lymphadenopathy associated virus (LAV) -
PT   used for prodn. of DNA, antigens and antibodies used in
PT   diagnosis of AIDS and pre-AIDS
PS   Claim 3; Fig 7A-7J; 72pp; English.
CC   LAV EL I (n80436) and LAV HA L (n80437) were isolated from the peripheral
CC   blood lymphocytes of patients. The different AIDS virus isolates
CC   are designated by 3 letters of the patients name. Stable probes including
CC   the DNA sequences can be used for detection of the new LAV viruses or
CC   related viruses or DNA proviruses in eg biological samples. The proteins
CC   or peptides can be used for detection of antibodies induced in vivo and
CC   present in biological fluids. The DNA can also be used for the expression
CC   of LAV viral antigens for the prodn. of a vaccine against LAV. The
CC   polypeptides can also be used for the prodn. of antibodies for the
CC   detection of proteins related to the LAV viruses, partic. for diagnosis
```

```
CC   of AIDS or pre-AIDS.
SQ   Sequence   9236 BP;   3360 A;   1642 C;   2190 G;   2044 T;

Initial Score    =    502  Optimized Score  =    502  Significance = 30.62
Residue Identity =    89%  Matches          =    502  Mismatches   =    57
Gaps             =      0  Conservative Substitutions               =     0

                                                    X       10        20
                                                    GGGGGACTGGAAGGGCTAATTC
                                                    ||||||||||||||||||||||
   CAATGACTTACAAAGAAGCTCTAGATCTCAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTT
  8630      8640      8650      8660      8670      8680      8690

        30        40        50        60        70        80        90
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
     ||| ||   |||||||| |||||||||||| ||| ||||| |||||||||||  ||||||||||||||| |
   GGTCCAAAAAGAGACAAGAGATCCTTGATCTTTGGGTCTACAACACACAAGGCATCTTCCCTGATTGGCAAA
8700      8710      8720      8730      8740      8750      8760      8770

      100       110       120       130       140       150       160
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
   |||||||||||||||||||| |||||||||||||| |||||||||||||||||| |||||||||||||||| |
   ACTACACACCAGGGCCAGGGATCAGATATCCACTAACCTTTGGATGGTGCTACGAGCTAGTACCAGTTGATC
        8780      8790      8800      8810      8820      8830      8840

    170       180       190       200       210       220       230
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
   || |  |||||||||| | || | ||||||||| || |||||||||||||||| |  ||| ||||||||||
   CACAGGAGGTAGAAGAAGACACTGAAGGAGAGACCAACAGCTTGTTACACCCTATATGCCAGCATGGAATGG
      8850      8860      8870      8880      8890      8900      8910

  240       250       260       270       280       290       300       310
   ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
   | ||||| |||||| |||||||| | ||||| ||| ||||| | ||||||||| | |||  ||||||||||
   AGGACCCGGAGAGACAAGTGTTAAAATGGAGATTTAACAGCAGACTAGCATTTGAGCACAAGGCCCGAGAGA
    8920      8930      8940      8950      8960      8970      8980

        320       330       340       350       360       370       380
   TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
   |||||||||||| || ||| ||||| ||||| |||||||| ||||||||||||||||||||||| |||||||||
   TGCATCCGGAGTTCTACAAAAACTGATGACACCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAG
  8990      9000      9010      9020      9030      9040      9050

      390       400       410       420       430       440       450
   GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
   |||||||||| ||||||||  |||||||||||||| | ||||||||||||||||||||||||||||||||||
   GGAGGCGTGGACTGGGCGGGACTGGGGAGTGGCTAACCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
9060      9070      9080      9090      9100      9110      9120      9130

     460       470       480       490       500       510       520
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
   |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
   TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAGCTAGGGAACCCACTGCTT
      9140      9150      9160      9170      9180      9190      9200

    530       540       550       560       570       580       590
   AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAG
   |||||||||||||||||||||||||||||||||
   AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAA
      9210      9220      9230     X

  600
   AGATCCCTCAG
```

```
> 0 <
0| |0  IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file railey-000-716.res made by shears on Mon 26 Apr 93 15:30:46-PDT.


Query sequence being compared:RAILEY-000-716.SEQ (1-696)
Number of sequences searched:              128494
Number of scores above cutoff:               4938

      Results of the initial comparison of RAILEY-000-716.SEQ (1-696) with:
   Data bank : EMBL-NEW 2, all entries
   Data bank : GenBank 75, all entries
   Data bank : GenBank-NEW 2, all entries
   Data bank : UEMBL 33_75, all entries

100000-
      -
N     -
U50000- *
M     -  *
B     -*
E     -
R     -
      -   *
0     -
F10000-
      -
S     -
E 5000-    *
Q     -
U     -
E     -
N     *
C     -     *
E     -
S 1000-
      -
      -
   500-
      -      *
      -
      -
      -
      -       *
      -
   100-
      -                                        *
      -        *
    50-
      -
      -
      -         *                                *  *
```



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	10
Optimized scores to save	0	Display context	50

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	30	30	12.19

Times:	CPU	Total Elapsed
	00:44:53.05	01:01:44.00

Number of residues: 154807074
Number of sequences searched: 128494
Number of scores above cutoff: 4938

Cut-off raised to 24.
Cut-off raised to 28.
Cut-off raised to 31.
Cut-off raised to 34.
Cut-off raised to 37.
Cut-off raised to 40.
Cut-off raised to 43.
Cut-off raised to 46.
Cut-off raised to 48.
Cut-off raised to 51.
Cut-off raised to 53.
Cut-off raised to 56.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame

```
                    **** 53 standard deviations above mean ****
  1. HIVPV22        Human immunodeficiency virus    9770   681   684  53.41   0
                    **** 52 standard deviations above mean ****
  2. HIVHXB2CG      Human immunodeficiency virus    9718   664   671  52.01   0
  3. REHTLV3        Human T-cell leukaemia type I   9748   664   671  52.01   0
  4. HIVH3CG        Human T-cell lymphotropic vir   9749   664   671  52.01   0
                    **** 51 standard deviations above mean ****
  5. HIVJRCSF       Human immunodeficiency virus    9540   652   652  51.03   0
                    **** 50 standard deviations above mean ****
  6. HIVNY5CG       Human immunodeficiency virus    9022   650   650  50.86   0
  7. HIVNL43        Human immunodeficiency virus    9709   645   645  50.45   0
  8. AIHTLV31       Human t-cell leukemia virus t    660   644   645  50.37   0
                    **** 49 standard deviations above mean ****
  9. REHIVXB2       Human T-lymphotropic virus ty    923   631   631  49.31   0
                    **** 48 standard deviations above mean ****
 10. REHIVXB3       Human T-lymphotropic virus ty    923   626   626  48.90   0
 11. HIVZ6          Human immunodeficiency virus    5159   626   626  48.90   0
 12. HIVZ2Z6        Human immunodeficiency virus    9081   626   626  48.90   0
                    **** 47 standard deviations above mean ****
 13. HIVSF2B13      Human immunodeficiency virus    3983   605   605  47.17   0
 14. HIVSF2B13      Human immunodeficiency virus    3983   605   605  47.17   0
                    **** 46 standard deviations above mean ****
 15. REHIVAT3       Human T-lymphotropic virus ty    917   598   598  46.60   0
 16. HIVIHB101      Human Immunodeficiency virus    9781   596   507  46.43   0
                    **** 44 standard deviations above mean ****
 17. HIVSFAAA       Human immunodeficiency virus    3954   574   603  44.63   0
                    **** 43 standard deviations above mean ****
 18. HIVMNCG        Human immunodeficiency virus    9738   563   639  43.73   0
 19. HIVBRUCG       Human immunodeficiency virus    9229   556   556  43.15   0
                    **** 42 standard deviations above mean ****
 20. REHIVC15       Human T-lymphotropic virus ty    769   550   550  42.66   0
 21. HL2ORF         Human T-cell lymphotropic vir    768   549   549  42.58   0
 22. HIVPCV12       Human immunodeficiency virus    2304   542   544  42.00   0
                    **** 41 standard deviations above mean ****
 23. HIVNE033       Human immunodeficiency virus     851   537   537  41.59   0
 24. HIVNE002       Human immunodeficiency virus     851   537   537  41.59   0
 25. HIVNE037       Human immunodeficiency virus     851   535   535  41.43   0
 26. HIVNE103       Human immunodeficiency virus     851   534   534  41.35   0
 27. HIVNE038       Human immunodeficiency virus     851   534   534  41.35   0
 28. HIVNE031       Human immunodeficiency virus     851   534   534  41.35   0
 29. HIVNE023       Human immunodeficiency virus     851   534   534  41.35   0
 30. HIVNE005       Human immunodeficiency virus     851   534   534  41.35   0
 31. HIVNE004       Human immunodeficiency virus     851   534   534  41.35   0
 32. HIVNE001       Human immunodeficiency virus     851   534   534  41.35   0
 33. HIVNE087       Human immunodeficiency virus     851   533   533  41.27   0
 34. HIVNE084       Human immunodeficiency virus     851   533   533  41.27   0
 35. HIVNE046       Human immunodeficiency virus     851   533   533  41.27   0
 36. HIVNE040       Human immunodeficiency virus     851   533   533  41.27   0
 37. HIVNE036       Human immunodeficiency virus     851   533   533  41.27   0
 38. HIVNE032       Human immunodeficiency virus     851   533   533  41.27   0
 39. HIVNE027       Human immunodeficiency virus     851   533   533  41.27   0
 40. HIVNE022       Human immunodeficiency virus     851   533   533  41.27   0


1. RAILEY-000-716.SEQ (1-696)
   HIVPV22      Human immunodeficiency virus type 1, isolate PV22,

 LOCUS       HIVPV22      9770 bp ss-RNA             VRL       15-MAR-1990
 DEFINITION  Human immunodeficiency virus type 1, isolate PV22, complete genome
             (H9/HTLV-III proviral DNA).
 ACCESSION   K02083
 KEYWORDS    TAR protein; acquired immune deficiency syndrome; complete genome;
             env protein; gag protein; long terminal repeat (LTR); pol protein;
             polyprotein; proviral gene; rev protein; reverse transcriptase;
             tat protein; trans-activator.
```

```
SOURCE      Human immunodeficiency virus type 1 (HIV-1), isolate PV22 (from
            H9-derived family), proviral DNA.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9770; 1 to 9770)
  AUTHORS   Muesing,M.A., Smith,D.H., Cabradilla,C.D., Benton,C.V., Kasky,L.A.
            and Capon,D.J.
  TITLE     Nucleic acid structure and expression of the human AIDS/
            lymphadenopathy retrovirus
  JOURNAL   Nature 313, 450-458 (1985)
  STANDARD  full automatic
REFERENCE   2  (sites)
  AUTHORS   van Beveren,C.P., Coffin,J. and Hughes,S.
  TITLE     Appendix B: HTLV-3/LAV genome
  JOURNAL   (in) Weiss,R., Teich,N., Varmus,H. and Coffin,J. (Eds.);
            RNA TUMOR VIRUSES, MOLECULAR BIOLOGY OF TUMOR VIRUSES, SECOND
            EDITION, 2 : SUPPLEMENTS AND APPENDIXES:  1106-1123,
            Cold Spring Harbor Laboratory, CSH, NY (1985)
  STANDARD  full automatic
REFERENCE   3  (bases 2111 to 2112)
  AUTHORS   Muesing,M.A.
  JOURNAL   Unpublished (1987) Whitehead Inst Cambridge, Mass
  STANDARD  full automatic
COMMENT     [1]  revised sequence, personal communication.
            [(in) Weiss,R., Teich,N., Varmus,H. and Coffin,J. (Eds.);RNA Tumor
            Viruses,Molecu]  review; bases 1 to 9769.
            [3]  revises [1],[(in) Weiss,R., Teich,N., Varmus,H. and Coffin,J.
            (Eds.);RNA Tumor Viruses,Molecu].

            This sequence for a H9/HTLV-III virus was determined from one
            complete proviral clone [1]. Additionally, several cDNA clones of
            the viral RNA were sequenced for comparison with the entire
            proviral sequence. The differences between cDNA and proviral DNA
            are extensive and are listed in the Sites Table as variations. The
            authors believe that the variations may be due in part to different
            strains in the H9/HTLV-III cell line, because it was established by
            infection with material from several AIDS patients.

            With the addition of g at 2111, gag cds and pol cds are very close
            to those of HXB2, BRU, and related HIV viruses.

            For details and other references pertaining to Sites and Features,
            see the HIV reference entry.
FEATURES             Location/Qualifiers
     cellular        1..9
                     /note="human cellular DNA"
     LTR             10..643
                     /note="5' LTR"
     repeat_region   463..560
                     /note="R repeat 5' copy"
     prim_transcript 464..9678
                     /note="genomic mRNA"
     prim_transcript 464..9678
                     /note="tat, rev, nef subgenomic mRNA"
     misc_feature    464
                     /note="numbered 1 in [1]"
     variation       510
                     /note="a in provirus; g in cDNA [1]"
     variation       575
                     /note="g in provirus; a in cDNA [1]"
     intron          753..5822
                     /note="tat, rev, nef subgenomic mRNA intron 1"
     variation       5716
                     /note="g in provirus; a in cDNA [1]"
     variation       5992
```

```
                /note="a in provirus; g in cDNA [1]"
variation       6007
                /note="c in provirus; t in cDNA [1]"
variation       6047
                /note="c in provirus; g in cDNA [1]"
variation       6051
                /note="c in provirus; a in cDNA [1]"
variation       6055..6057
                /note="agg in provirus; gaa in cDNA [1]"
intron          6091..8420
                /note="tat cds intron 2"
intron          6091..8420
                /note="rev cds intron 2"
intron          6091..8420
                /note="tat, rev, nef subgenomic mRNA intron 2"
variation       6108
                /note="t in provirus; c in cDNA [1]"
variation       6120
                /note="a in provirus; c in cDNA [1]"
variation       6125..6126
                /note="gc in provirus; gtaac in cDNA [1]"
variation       6136
                /note="a in provirus; c in cDNA [1]"
variation       6235
                /note="t in provirus; a in cDNA [1]"
variation       6352
                /note="g in provirus; a in cDNA [1]"
variation       6760
                /note="t in provirus; a in cDNA [1]"
variation       7090
                /note="c in provirus; t in cDNA [1]"
variation       7100
                /note="a in provirus; g in cDNA [1]"
variation       7134..7135
                /note="ca in provirus; ac in cDNA [1]"
variation       7183..7184
                /note="gt in provirus; aa in cDNA [1]"
variation       7199
                /note="a in provirus; g in cDNA [1]"
variation       7284..7285
                /note="aa in provirus; gc in cDNA [1]"
variation       7303
                /note="a in provirus; c in cDNA [1]"
variation       7511
                /note="a in provirus [1]; c in cDNA [1]"
variation       7533
                /note="t in provirus [1]; a in cDNA [1]"
variation       7586
                /note="c in provirus [1]; t in cDNA [1]"
variation       7648
                /note="a in provirus [1]; g in cDNA [1]"
variation       8139
                /note="a in provirus; c in cDNA [1]"
variation       8143
                /note="t in provirus; c in cDNA [1]"
variation       8222
                /note="g in provirus; a in cDNA [1]"
variation       8269
                /note="a in provirus [1]; g in cDNA [1]"
variation       8285
                /note="g in provirus [1]; t in cDNA [1]"
variation       8376
                /note="a in provirus [1]; g in cDNA [1]"
variation       8381
                /note="a in provirus [1]; g in cDNA [1]"
variation       8476
```

```
                /note="a in provirus [1]; g in cDNA [1]"
variation       8869
                /note="a in provirus [1]; g in cDNA [1]"
variation       8979
                /note="c in provirus; t in cDNA [1]"
variation       8990
                /note="a in provirus; c in cDNA [1]"
variation       8999
                /note="c in provirus [1]; a in cDNA [1]"
variation       9031
                /note="a in provirus [1]; g in cDNA [1]"
LTR             9128..9761
                /note="3' LTR"
variation       9291
                /note="t in provirus [1]; g in cDNA [1]"
variation       9295
                /note="g in provirus [1]; t in cDNA [1]"
variation       9303
                /note="g in provirus [1]; a in cDNA [1]"
variation       9548
                /note="g in provirus [1]; c in cDNA [1]"
repeat_region   9581..9678
                /note="R repeat 3' copy"
polyA_signal    9654..9659
                /note="mRNA polyadenylation signal"
cellular        9762..9770
                /note="human cellular DNA"
CDS             join(5876..6090,8421..8466)
                /note="tat protein"
                /codon_start=1
                /translation="MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALG
                ISYGRKKRRQRRRPPQGSQTHQVSLSKQPTSQPRGDPTGPKE"
CDS             join(6015..6090,8421..8695)
                /note="rev protein"
                /codon_start=1
                /translation="MAGRSGDSDEDLLKAVRLIKFLYQSNPPPNPEGTRQARRNRRRR
                WRERQRQIHSISERILSTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTILESGAKE"
CDS             799..2337
                /note="gag polyprotein precursor"
                /codon_start=1
                /translation="MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERF
                AVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALD
                KIEEEQNKSKKKAQQAAADTGHSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVE
                EKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPV
                HAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRM
                YSPTSILDIRQGPKEPFRDYLDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTIL
                KALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNTATIMMQRGNFRNQRKMVKC
                FNCGKEGHTARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFLQ
                SRPEPTAPPFLQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGND
                PSSQ"
CDS             2094..5141
                /partial
                /note="pol polyprotein;  (NH2-terminus uncertain)"
                /codon_start=1
                /translation="FFREDLAFLQGKAREFSSEQTRANSPTISSEQTRANSPTRREL
                VWGRDNNSPSEAGADRQGTVSFNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLE
                EMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQ
                IGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIG
                PENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL
                DVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKI
                LEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEP
                PFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLL
                RGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQI
                YQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWE
                TWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETRLGKAGY
```

```
                     LTNKGRQKVVPLTNTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSES
                     ELVNQIIEQLIKKQKVYLAWVPAHKGIGGNEQVDKLVSAGIRKILFLDGIDKAQDEHE
                     KYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGK
                     VILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTSATVKAA
                     CWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKR
                     KGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRNPLWKGPAKLLWKG
                     EGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED"
     CDS             5086..5664
                     /note="vif protein"
                     /codon_start=1
                     /translation="MENRWQVMIVWQVDRMRIRTWKSLVKHHMYVSGKARGWFYRHHY
                     ESPHPRISSEVHIPLGDARLVITTYWGLHTGERDWHLGQGVSIEWRKKRYSTQVDPEL
                     ADQLIHLYYFDCFSDSAIRKALLGHIVSPRCEYQAGHNKVGSLQYLALAALITPKKIK
                     PPLPSVTKLTEDRWNKPQKTKGHRGSHTMNGH"
     CDS             5604..5840
                     /note="vpr protein"
                     /codon_start=1
                     /translation="MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQ
                     HIYETYGDTWAGVEAIIRILQQLLFIHFQNWVST"
     CDS             6107..6352
                     /note="vpu protein"
                     /codon_start=1
                     /translation="MQPIQIAIVALVVAIIIAILVWSIVIIEYRKILRQRKIDRLIDR
                     LIERAEDSGNESEGEISALVEMGVEMGHHAPWDVDDL"
     CDS             6267..8837
                     /note="envelope polyprotein"
                     /codon_start=1
                     /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                     WKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVE
                     QMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
                     ISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYTLTSCNTSVITQACPKVSFEPIPIHYC
                     APAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSAN
                     FTDNAKTIIVQLNQSVEINCTRPNNNTRKSIRIQRGPGRAFVTIGKIGNMRQAHCNIS
                     RAKWNNTLKQIDSKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
                     STWFNSTWSTEGSNNTEGSDTITLPCRIKQFINMWQEVGKAMYAPPISGQIRCSSNIT
                     GLLLTRDGGNNNNESEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQR
                     EKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLL
                     QLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWN
                     NMTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWANLWNWLNITNWLWYI
                     KLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEDGER
                     DRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWN
                     LLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGAYRAIRHIPRRIRQGLERILL
                     "
     CDS             8839..9459
                     /note="nef protein"
                     /codon_start=1
                     /translation="MGGKWSKSSVIGWPAVRERMRRAEPAADGVGAASRDLEKHGAIT
                     SSNTAANNAACAWLEAQEEEKVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIH
                     SQRRQDILDLWIYHTQGYFPDWQNYTPGPGIRYPLTFGWCYKLVPVEPDKVEEANKGE
                     NTSLLHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFKNC"
 BASE COUNT     3436 a   1786 c   2376 g   2172 t
 ORIGIN      482 bp upstream of BglII site.

Initial Score    =    681  Optimized Score  =    684  Significance = 53.41
Residue Identity =    98%  Matches          =    685  Mismatches   =     9
Gaps             =      2  Conservative Substitutions               =     0

       X       10        20        30        40        50        60
       GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACAC
       | |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TGTAGTGGG--TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACAC
       X     10        20        30        40        50        60        70

   70        80        90       100       110       120       130       140
   ACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATG
   |||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||
```

```
ACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGACCAGGGATCAGATATCCACTGACCTTTGGATG
        80        90       100       110       120       130       140

       150       160       170       180       190       200       210
GTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTT
|||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
GTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAACAAAGGAGAGAACACCAGCTTGTT
     150       160       170       180       190       200       210

     220       230       240       250       260       270       280
ACACCCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCT
|||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
ACACCCTGTGAGCCTGCATGGAATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCT
   220       230       240       250       260       270       280

   290       300       310       320       330       340       350
AGCATTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAA
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
AGCATTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGATATCGAGCTTGCTACAA
 290       300       310       320       330       340       350

 360       370       380       390       400       410       420
GGGACTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATG
|||||||||||||||| |||||||||||||||||||||||||||| |||||||||||||||||||||||||||
GGGACTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATG
360       370       380       390       400       410       420       430

430       440       450       460       470       480       490       500
CTGCATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCT
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
CTGCATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCT
       440       450       460       470       480       490       500

       510       520       530       540       550       560       570
CTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCC
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCC
     510       520       530       540       550       560       570

     580       590       600       610       620       630       640
GTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTG
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTG
   580       590       600       610       620       630       640

   650       660       670       680       690     X
GCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
|||||||||||||||||||||||||||||||||||||||||||||||||||
GCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGA
 650       660       670       680       690       700       710

AGCGCGCACGGCAAGAGGCGAGGGGCGGCG
720       730       740
```

2. RAILEY-000-716.SE0 (1-696)
   HIVHXB2CG Human immunodeficiency virus type 1 (HXB2), comple

```
LOCUS       HIVHXB2CG    9718 bp ss-RNA             VRL       14-JAN-1992
DEFINITION  Human immunodeficiency virus type 1 (HXB2), complete genome;
            HIV1/HTLV-III/LAV reference genome.
ACCESSION   K03455
KEYWORDS    TAR protein; acquired immune deficiency syndrome; complete genome;
            env protein; gag protein; long terminal repeat (LTR); pol protein;
            polyprotein; proviral gene; reverse transcriptase; trans-activator.
SOURCE      HTLV-III/LAV (isolate HXB2) proviral DNA.
```

ORGANISM Human immunodeficiency virus type 1
Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
Retroviridae; Lentivirinae.
REFERENCE 1 (sites)
AUTHORS Rosen,C.A., Sodroski,J.G. and Haseltine,W.A.
TITLE The location of cis-acting regulatory sequences in the human T cell
lymphotropic virus type III (HTLV-III/LAV) long terminal repeat
JOURNAL Cell 41, 813-823 (1985)
STANDARD full automatic
REFERENCE 2 (bases 9577 to 9718; 493 to 674)
AUTHORS Wong-Staal,F., Gallo,R.C., Chang,N.T., Ghrayeb,J., Papas,T.S.,
Lautenberger,J.A., Pearson,M.L., Petteway,S.R.Jr., Ivanoff,L.,
Baumeister,K., Whitehorn,E.A., Rafalski,J.A., Doran,E.R.,
Josephs,S.J., Starcich,B., Livak,K.J., Patarca,R., Haseltine,W.A.
and Ratner,L.
TITLE Complete nucleotide sequence of the AIDS virus, HTLV-III
JOURNAL Nature 313, 277-284 (1985)
STANDARD full automatic
REFERENCE 3 (sites)
AUTHORS van Beveren,C.P., Coffin,J. and Hughes,S.
TITLE Appendix B: HTLV-3/LAV genome
JOURNAL (in) Weiss,R., Teich,N., Varmus,H. and Coffin,J. (Eds.);
RNA TUMOR VIRUSES, SECOND EDITION, 2: 1102-1123,
Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
STANDARD full automatic
REFERENCE 4 (bases 1 to 653)
AUTHORS Starcich,B., Ratner,L., Josephs,S.F., Okamato,T., Gallo,R.C. and
Wong-Staal,F.
TITLE Characterization of long terminal repeat sequences of HTLV-III
JOURNAL Science 227, 538-540 (1985)
STANDARD full automatic
REFERENCE 5 (sites)
AUTHORS Allan,J.S., Coligan,J.E., Barin,F., McLane,M.F., Sodroski,J.G.,
Rosen,C.A., Haseltine,W.A., Lee,T.H. and Essex,M.
TITLE Major glycoprotein antigens that induce antibodies in AIDS patients
are encoded by HTLV-III
JOURNAL Science 228, 1091-1094 (1985)
STANDARD full automatic
REFERENCE 6 (sites)
AUTHORS Arya,S.K., Guo,C., Josephs,S.F. and Wong-Staal,F.
TITLE Trans-activator gene of human T-lymphotropic virus type III
(HTLV-III)
JOURNAL Science 229, 69-73 (1985)
STANDARD full automatic
REFERENCE 7 (sites)
AUTHORS Sodroski,J., Patarca,R., Rosen,C., Wong-Staal,F. and Haseltine,W.A.
TITLE Location of the trans-activating region on the genome of human
T-cell lymphotropic virus type III
JOURNAL Science 229, 74-77 (1985)
STANDARD full automatic
REFERENCE 8 (sites)
AUTHORS Rabson,A.B., Daugherty,D.F., Venkatesan,S., Boulukos,K.E.,
Benn,S.I., Folks,T.M., Feorino,P. and Martin,M.
TITLE Transcription of novel open reading frames of AIDS retrovirus
during infection of lymphocytes
JOURNAL Science 229, 1388-1390 (1985)
STANDARD full automatic
REFERENCE 9 (sites)
AUTHORS Allan,J.S., Coligan,J.E., Lee,T.-H., McLane,M.F., Kanki,P.J.,
Groopman,J.E. and Essex,M.
TITLE A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS
patients
JOURNAL Science 230, 810-813 (1985)
STANDARD full automatic
REFERENCE 10 (sites)
AUTHORS Dauton,A.I., Sodroski,J.G., Rosen,C.A., Goh,W.C. and Haseltine,W.A.

TITLE The trans-activator gene of the human T cell lymphotropic virus type III is required for replication
JOURNAL Cell 44, 941-947 (1986)
STANDARD full automatic
REFERENCE 11 (sites)
AUTHORS Starcich,B.R., Hahn,B.H., Shaw,G.M., McNeely,P.D., Modrow,S., Wolf,H., Parks,E.S., Parks,W.P., Josephs,S.F., Gallo,R.C. and Wong-Staal,F.
TITLE Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS
JOURNAL Cell 45, 637-648 (1986)
STANDARD full automatic
REFERENCE 12 (sites)
AUTHORS Feinberg,M.B., Jarret,R.F., Aldovini,A., Gallo,R.C. and Wong-Staal,F.
TITLE HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA
JOURNAL Cell 46, 807-817 (1986)
STANDARD full automatic
REFERENCE 13 (sites)
AUTHORS Terwilliger,E., Sodroski,J.G., Rosen,C.A. and Haseltine,W.A.
TITLE Effects of mutations within the 3' orf open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenecity
JOURNAL J. Virol. 60, 754-760 (1986)
STANDARD full automatic
REFERENCE 14 (sites)
AUTHORS Lightfoote,M.M., Coligan,J.E., Folks,T.M., Fauci,A.S., Martin,M.A. and Venkatesan,S.
TITLE Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus
JOURNAL J. Virol. 60, 771-775 (1986)
STANDARD full automatic
REFERENCE 15 (sites)
AUTHORS Rosen,C.A., Sodroski,J.G., Goh,W.C., Dayton,A.I., Lippke,J.A. and Haseltine,W.A.
TITLE Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III
JOURNAL Nature 319, 555-559 (1986)
STANDARD full automatic
REFERENCE 16 (sites)
AUTHORS Sodroski,J., Goh,W.C., Rosen,C., Dayton,A.I., Terwilliger,E. and Haseltine,W.A.
TITLE A second post-transcriptional trans-activator gene required for HTLV-III replication
JOURNAL Nature 321, 412-417 (1986)
STANDARD full automatic
REFERENCE 17 (sites)
AUTHORS Arya,S.K. and Gallo,R.C.
TITLE Three novel genes of human T-lymphotropic virus type III: Immune reactivity of their products with sera from acquired immune deficiency syndrome patients
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83, 2209-2213 (1986)
STANDARD full automatic
REFERENCE 18 (sites)
AUTHORS Willey,R., Rutledge,R.A., Dias,S., Folks,T., Theodore,T., Buckler,C.E. and Martin,M.A.
TITLE Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83, 5038-5042 (1986)
STANDARD full automatic
REFERENCE 19 (sites)
AUTHORS di Marzo Veronese,F., Copeland,T.D., DeVico,A.L., Rahman,R., Oroszlan,S., Gallo,R.C. and Sarngadharan,M.G.

TITLE Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV
JOURNAL Science 231, 1289-1291 (1986)
STANDARD full automatic
REFERENCE 20 (sites)
AUTHORS Lee,T.-H., Coligan,J.E., Allan,J.S., McLane,M.F., Groopman,J.E. and Essex,M.
TITLE A new HTLV-III/LAV protein encoded by a gene found in cytopathic retroviruses
JOURNAL Science 231, 1546-1549 (1986)
STANDARD full automatic
REFERENCE 21 (sites)
AUTHORS Sodroski,J., Goh,W.C., Rosen,C., Tartar,A., Portetelle,D., Burny,A. and Haseltine,W.A.
TITLE Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions
JOURNAL Science 231, 1549-1553 (1986)
STANDARD full automatic
REFERENCE 22 (sites)
AUTHORS Kan,N.C., Franchini,G., Wong-Staal,F., DuBois,G.C., Robey,W.G., Lautenberger,J.A. and Papas,T.S.
TITLE Identification of HTLV-III/LAV sor gene product and detection of antibodies in human sera
JOURNAL Science 231, 1553-1555 (1986)
STANDARD full automatic
REFERENCE 23 (sites)
AUTHORS Kramer,R.A., Schaber,M.D., Skalka,A.M., Ganguly,K., Wong-Staal,F. and Reddy,P.E.
TITLE HTLV-III gag protein is processed in yeast cells by the virus pol-protease
JOURNAL Science 231, 1580-1584 (1986)
STANDARD full automatic
REFERENCE 24 (sites)
AUTHORS Jones,K.A., Kadonaga,J.T., Luciw,P.A. and Tjian,R.
TITLE Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1
JOURNAL Science 232, 755-759 (1986)
STANDARD full automatic
REFERENCE 25 (bases 8761 to 9060)
AUTHORS Fisher,A.G., Ratner,L., Mitsuya,H., Marselle,L.M., Harper,M.E., Broder,S., Gallo,R.C. and Wong-Staal,F.
TITLE Infectious mutants of HTLV-III with changes in the 3' region and markedly reduced cytopathic effects
JOURNAL Science 233, 655-659 (1986)
STANDARD full automatic
REFERENCE 26 (sites)
AUTHORS Wright,C.M., Felber,B.K., Paskalis,H. and Pavlakis,G.N.
TITLE Expression and characterization of the trans-activator of HTLV-III/LAV virus
JOURNAL Science 234, 988-992 (1986)
STANDARD full automatic
REFERENCE 27 (bases 5611 to 5611)
AUTHORS Ratner,L.
JOURNAL Unpublished (1987) Washington U Med School, St. Louis, MO
STANDARD full automatic
REFERENCE 28 (sites)
AUTHORS Wong-Staal,F., Chanda,P.K. and Ghrayeb,J.
TITLE Human immunodeficiency virus: the eighth gene
JOURNAL AIDS Res. Hum. Retroviruses 3, 33-39 (1987)
STANDARD full automatic
REFERENCE 29 (sites)
AUTHORS Patarca,R., Heath,C., Goldenberg,G.J., Rosen,C.A., Sodroski,J.G., Haseltine,W.A. and Hansen,U.M.
TITLE Transcription directed by the HIV long terminal repeat in vitro
JOURNAL AIDS Res. Hum. Retroviruses 3, 41-55 (1987)
STANDARD full automatic

REFERENCE 30 (bases 1 to 9635; 1 to 9635)
AUTHORS Ratner,L., Fisher,A., Jagodzinski,L.L., Mitsuya,H., Liou,R.-S., Gallo,R.C. and Wong-Staal,F.
TITLE Complete nucleotide sequences of functional clones of the AIDS virus
JOURNAL AIDS Res. Hum. Retroviruses 3, 57-69 (1987)
STANDARD full automatic
REFERENCE 31 (sites)
AUTHORS Muesing,M.A., Smith,D.H. and Capon,D.J.
TITLE Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein
JOURNAL Cell 48, 691-701 (1987)
STANDARD full automatic
REFERENCE 32 (sites)
AUTHORS Modrow,S., Hahn,B.H., Shaw,G.M., Gallo,R.C., Wong-Staal,F. and Wolf,H.
TITLE Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: Prediction of antigenic epitopes in conserved and variable regions
JOURNAL J. Virol. 61, 570-578 (1987)
STANDARD full automatic
REFERENCE 33 (sites)
AUTHORS Goh,W.C., Sodroski,J.G., Rosen,C.A. and Haseltine,W.A.
TITLE Expression of the art gene protein of human T-lymphotropic virus type III (HTLV-III/LAV) in bacteria
JOURNAL J. Virol. 61, 633-637 (1987)
STANDARD full automatic
REFERENCE 34 (sites)
AUTHORS Nabel,G. and Baltimore,D.
TITLE An inducible transcription factor activates expression of human immunodeficiency virus in T cells
JOURNAL Nature 326, 711-713 (1987)
STANDARD full automatic
REFERENCE 35 (sites)
AUTHORS Fisher,A.G., Ensoli,B., Ivanoff,L., Chamberlain,M., Petteway,S., Ratner,L., Gallo,R.C. and Wong-Staal,F.
TITLE The sor gene of hiv-1 is required for efficient virus transmission in vitro
JOURNAL Science 237, 888-893 (1987)
STANDARD full automatic
REFERENCE 36 (sites)
AUTHORS Ido,E., Han,H.-p., Kezdy,F.J. and Tang,J.
TITLE Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S
JOURNAL J. Biol. Chem. 266, 24359-24366 (1991)
STANDARD full automatic
COMMENT [6] sites; tat mRNA and other transcript boundaries.
[7] sites; tat mRNA.
[8] sites; mRNA splice sites.
[9] sites; 27K antigen cds.
[5] sites; gp160 and gp120 coding sequences.
[1] sites; regulatory sequences in the LTR.
[(in) Weiss,R., Teich,N., Varmus,H. and Coffin,J. (Eds.);RNA Tumor Viruses, Secon] review; bases 1 to 9718.
[15] sites; trans-activator function and TAR sequence.
[19] sites; pol coding sequence.
[22] sites; 23K sor gene product.
[23] sites; pol NH2-terminal region.
[20] sites; sor 23K protein.
[21] sites; sor 23K protein.
[24] sites; Sp1 binding sites in the promoter region.
[17] sites; acceptor and donor splice sites for tat and 27K. [10] sites; deletion mutants in the tat gene.
[18] sites; env gene conserved/varable regions; separate entries.
[16] sites; trs cds boundaries.
[12] sites; trs cds boundaries.

```
            [11]  sites; env gene conserved/variable regions; separate entries.
            [26]  sites; tar or transactivator target.
            [13]  sites; 3' orf mutations.
            [14]  sites; pol p34 terminus.
            [31]  sites; promoter, TAR, tat-III mutants.
            [32]  sites; envelope protein epitopes.
            [33]  sites; trs/art protein.
            [34]  sites; inducible enhancer element.
            [27]  revises [30].
            [29]  sites; long terminal repeat.
            [28]  sites; R orf.
            [35]  sites; sor.

            Sequence for [25] kindly provided in computer-readable form by
            L.Ratner, 19-AUG-1986.

            The HXB2 sequence is being used as a reference genome for all the
            HIV entries because it has been derived from a demonstrably
            infectious clone.  Hence not all of the 'sites' references above
            were concerned with this isolate.
FEATURES             Location/Qualifiers
     exon            <5830..6044
                     /number=2
                     /note="tat protein, (first expressed exon)"
     exon            <5969..6044
                     /number=2
                     /note="trs protein, (first expressed exon)"
     LTR             1..634
                     /note="5' LTR"
     repeat_region   454..551
                     /note="R repeat 5' copy"
     mRNA            455..9635
                     /note="HXB2 genomic mRNA"
     prim_transcript 455..9635
                     /note="tat, trs, 27K subgenomic mRNA"
     intron          743..5776
                     /note="tat,trs, 27K mRNA intron 1"
     intron          6045..8377
                     /note="tat intron 1"
     intron          6045..8377
                     /note="trs intron 2"
     intron          6045..8377
                     /note="27K mRNA intron 2"
     intron          6045..8377
                     /note="tat, trs intron 2"
     exon            8378..>8423
                     /number=3
                     /note="tat protein"
     exon            8378..>8652
                     /number=3
                     /note="trs protein"
     LTR             9085..9718
                     /note="3' LTR"
     repeat_region   9539..9635
                     /note="R repeat 3' copy"
     polyA_signal    9611..9616
                     /note="HXB2 mRNA polyadenyation signal"
     CDS             join(5830..6044,8378..8423)
                     /note="tat protein"
                     /codon_start=1
                     /translation="MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALG
                     ISYGRKKRRQRRRAHQNSQTHQASLSKQPTSQSRGDPTGPKE"
     CDS             join(5969..6044,8378..8652)
                     /note="trs protein"
                     /codon_start=1
                     /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARRNRRRR
```

```
                WRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTVLESGTKE"
CDS             789..2291
                /note="gag polyprotein"
                /codon_start=1
                /translation="MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERF
                AVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALD
                KIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVE
                EKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPV
                HAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRM
                YSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTIL
                KALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKIVKC
                FNCGKEGHTARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFLQ
                SRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQ"
CDS             2357..5095
                /partial
                /note="pol polyprotein;  (NH2-terminus uncertain)"
                /codon_start=1
                /translation="MSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGP
                TPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEIC
                TEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPH
                PAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWK
                GSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRW
                GLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQI
                YPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAE
                IQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKT
                PKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDG
                AANRETKLGKAGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYA
                LGIIQAQPDQSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL
                FLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSP
                GIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIHTD
                NGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKT
                AVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRN
                SLWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED"
CDS             5040..5618
                /note="sor 23K protein"
                /codon_start=1
                /translation="MENRWQVMIVWQVDRMRIRTWKSLVKHHMYVSGKARGWFYRHHY
                ESPHPRISSEVHIPLGDARLVITTYWGLHTGERDWHLGQGVSIEWRKKRYSTQVDPEL
                ADQLIHLYYFDCFSDSAIRKALLGHIVSPRCEYQAGHNKVGSLQYLALAALITPKKIK
                PPLPSVTKLTEDRWNKPQKTKGHRGSHTMNGH"
CDS             5558..5794
                /note="R;  (ORF) protein"
                /codon_start=1
                /translation="MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQ
                HIYETYGDTWAGVEAIIRILQQLLFIHFQNWVST"
CDS             6224..8794
                /note="envelope polyprotein"
                /codon_start=1
                /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
                WKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFDMWKNDMVE
                QMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
                ISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYSLTSCNTSVITQACPKVSFEPIPIHYC
                APAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVN
                FTDNAKTIIVQLNTSVEINCTRPNNNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNIS
                RAKWNNTLKQIDSKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
                STWFNSTWSTEGSNNTEGSDTITLPCRIKQIINMWQKVGKAMYAPPISGQIRCSSNIT
                GLLLTRDGGNSNNESEIFRLGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQR
                EKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLL
                QLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWN
                HTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYI
                KLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPIPRGPDRPEGIEEEGGER
                DRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWN
                LLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL
                "
CDS             8796..9167
```

/note="27K protein; [redacted] termination)"
/codon_start=1
/translation="MGGKWSKSSVIGWLTVRERMRRAEPAADGVGAASRDLEKHGAIT
SSNTAATNAACAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIH
SQRRQDILDLWIYHTQGYFPD"
BASE COUNT 3411 a 1773 c 2370 g 2164 t
ORIGIN 435 bp upstream of PvuII site; 5' end of proviral genome.

Initial Score = 664 Optimized Score = 671 Significance = 52.01
Residue Identity = 97% Matches = 673 Mismatches = 13
Gaps = 3 Conservative Substitutions = 0

```
          10        20        30        40        50        60        70
   GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          X       10        20        30        40        50        60

        80        90       100       110       120       130       140
   GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
   |||||||||||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
   GGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
     70        80        90       100       110       120       130

      150       160       170       180       190       200       210
   TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
   |||||||||||||||||||||||||| ||| ||||||| ||||| ||||||||||||||||||||||||||||
   TACAAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAAGCCAACAAAGGAGAGAACACCAGCTTGTTACAC
   140       150       160       170       180       190       200

    220       230       240       250       260       270       280
   CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
   |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
   CCTGTGAGCCTGCATGGAATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
 210       220       230       240       250       260       270       280

  290       300       310       320       330       340       350       360
   TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
   ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
         290       300       310       320       330       340       350

        370       380       390       400       410       420       430
   CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
   |||||||||||| |||||||||||||||||||||||||| |||||||||||||||||||||||| ||||
   CTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGC
      360       370       380       390       400       410       420

       440       450       460       470       480       490       500
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
   |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
     430       440       450       460       470       480       490

     510       520       530       540       550       560       570
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   500       510       520       530       540       550       560

    580       590       600       610       620       630       640
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
 570       580       590       600       610       620       630       640
```

```
  650       660       670       680       690     X
   CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
   |||||||||||| ||||||||||||||||||||   ||||||||||||
   CCGAACAGGGACCTGAAAGCGAAAGGGAAACCA---GAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
         650       660       670          680     X 690       700       710

   CCCGCACGGCAAGAGGCGAGGGGCGG
          720       730
```

3. RAILEY-000-716.SEQ (1-696)
   REHTLV3   Human T-cell leukaemia type III (HTLV-III) provira

```
LOCUS       REHTLV3      9748 bp    RNA             VRL       08-MAY-1992
DEFINITION  Human T-cell leukaemia type III (HTLV-III) proviral genome (AIDS
            virus for acquired immune deficiency syndrome)
ACCESSION   X01762
KEYWORDS    acquired immune deficiency syndrome; direct repeat; endonuclease;
            glycoprotein; inverted repeat; protease; provirus;
            reverse transcriptase; terminal repeat.
SOURCE      Human immunodeficiency virus type 1
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA viruses;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9748)
  AUTHORS   Wong-staal,F., Gallo,R.C., Chang,N.T., Ghrayeb,J., Papas,T.S.,
            Lautenberger,J.A., Pearson,M.L., Petteway,S.R.Jr., Ivanoff,L.,
            Baumeister,K., Whitehorn,E.A., Rafalski,J.A., Doran,E.R.,
            Josephs,S.J., Starcich,B., Livak,K.J., Patarca,R., Haseltine,W. and
            Ratner,L.
  TITLE     Complete nucleotide sequence of the AIDS virus, HTLV-III
  JOURNAL   Nature 313, 277-284 (1985)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 9748)
  AUTHORS   Muesing,M.A., Smith,D.H., Cabradilla,C.D., Benton,C.V., Kasky,L.A.
            and Capon,D.J.
  TITLE     Nucleic acid structure and expression of the human AIDS/
            lymphadenopathy retrovirus
  JOURNAL   Nature 313, 450-458 (1985)
  STANDARD  full automatic
FEATURES             Location/Qualifiers
     misc_feature    1..634
                     /note="long terminal repeat"
     repeat_unit     1..2
                     /note="inverted repeat"
     promoter        427..430
                     /note="TATA-box"
     misc_feature    453
                     /note="U3 region"
     misc_feature    454..551
                     /note="R region"
     misc_RNA        454
                     /note="cap site"
     misc_feature    552..634
                     /note="U5 region"
     repeat_unit     633..634
                     /note="inverted repeat"
     misc_feature    635..653
                     /note="tRNA binding site (tRNA-Lys)"
     repeat_region   1968..2002
                     /note="direct repeat"
     repeat_region   2031..2065
                     /note="direct repeat"
     repeat_region   2128..2163
                     /note="direct repeat"
     repeat_region   2164..2176
```

```
                /note="direct repeat"
misc_feature    7786..7787
                /note="put.peptide cleavage site"
misc_feature    9098..9103
                /note="poly purine stretch"
repeat_region   9115..9748
                /note="long terminal repeat"
misc_feature    9115..9567
                /note="U3 region"
misc_feature    9568..9665
                /note="R region"
misc_feature    9641..9646
                /note="polyadenylation signal"
misc_feature    9666..9748
                /note="U5 region"
repeat_unit     9747..9748
                /note="inverted repeat"
CDS             787..2321
                /note="gag precursor polypeptide"
CDS             1183..2321
                /note="gag p24 and gag p15 for major capsid protein and
                for put. retroviral nucleic acid binding protein
                (NBP)(ref.2) (boundaries not defined)"
CDS             787..1182
                /note="gag p17"
                /codon_start=1
                /translation="MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERF
                AVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALD
                KIEEEQNKSKKKAQQAAADTGHSSQVSQNY"
CDS             2081..5125
                /note="pol precursor polypeptides put. protease at 5'
                terminus reverse transcriptase put. endonuclease at 3'
                terminus"
                /codon_start=1
                /translation="FFREDLAFLQGKAREFSSEQTRANSPTISSEQTRANSPTRRELQ
                VWGRDNNSPSEAGADRQGTVSFNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLE
                EMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQ
                IGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIG
                PENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL
                DVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKI
                LEPFKKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEP
                PFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLL
                RGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQI
                YQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWE
                TWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGY
                VTNKGRQKVVPLTNTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDKSES
                ELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKILFLDGIDKAQDEHE
                KYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGK
                VILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTSATVKAA
                CWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKR
                KGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRNPLWKGPAKLLWKG
                EGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED"
CDS             5040..5648
                /note="SOR short open reading frame pot. vestigial env
                gene"
                /codon_start=1
                /translation="CQEEKQRSLGIMENRWQVMIVWQVDRMRIRTWKSLVKHHMYVSG
                KARGWFYRHHYESPHPRISSEVHIPLGDARLVITTYWGLHTGERDWHLGQGVSIEWRK
                KRYSTQVDPELADQLIHLYYFDCFSDSAIRKALLGHIVSPRCEYQAGHNKVGSLQYLA
                LAALITPKKIKPPLPSVTKLTEDRWNKPQKTKGHRGSHTMNGH"
CDS             6323..8821
                /note="env-lor precursor polypeptide"
                /codon_start=1
                /translation="MLMICSATEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHN
                VWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPL
                CVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFNISTSIRGKVQKEYAFFYKLDIIP
```

```
                     IDNDTTSYTLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGPCTN
                     VSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSANFTDNAKTIIVQLNQSVEINCTRP
                     NNNTRKSIRIQRGPGRAFVTIGKIGNMRQAHCNISRAKWNNTLKQIDSKLREQFGNNK
                     TIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTKGSNNTEGSDTIT
                     LPCRIKQIINMWQEVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESEIFRPGGG
                     DMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGIGALFLGFLGAAGSTMG
                     AASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKD
                     QQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWNNMTWMEWDREINNYTSLIHSLIE
                     ESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLSVV
                     NRVRQGYSPLSFQTHLPIPRGPDRPEGIEEEGGERDRDRSIRLVNGSLALIWDDLRSL
                     CLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLLQYWSQELKNSAVSLLNATAIA
                     VAEGTDRVIEVVQGAYRAIRHIPRRIRQGLERILL"
     CDS             6323..8821
                     /note="envelope glycoprotein"
                     /codon_start=1
                     /translation="MLMICSATEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHN
                     VWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPL
                     CVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFNISTSIRGKVQKEYAFFYKLDIIP
                     IDNDTTSYTLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGPCTN
                     VSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSANFTDNAKTIIVQLNQSVEINCTRP
                     NNNTRKSIRIQRGPGRAFVTIGKIGNMRQAHCNISRAKWNNTLKQIDSKLREQFGNNK
                     TIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTKGSNNTEGSDTIT
                     LPCRIKQIINMWQEVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESEIFRPGGG
                     DMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGIGALFLGFLGAAGSTMG
                     AASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKD
                     QQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWNNMTWMEWDREINNYTSLIHSLIE
                     ESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLSVV
                     NRVRQGYSPLSFQTHLPIPRGPDRPEGIEEEGGERDRDRSIRLVNGSLALIWDDLRSL
                     CLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLLQYWSQELKNSAVSLLNATAIA
                     VAEGTDRVIEVVQGAYRAIRHIPRRIRQGLERILL"
     CDS             7787..8821
                     /note="put.lor transmembrane protein"
                     /codon_start=1
                     /translation="AVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNL
                     LRAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNAS
                     WSNKSLEQIWNNMTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWN
                     WFNITNWLWYIKLFIMIVGGLVGLRIVFAVLSVVNRVRQGYSPLSFQTHLPIPRGPDR
                     PEGIEEEGGERDRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGR
                     RGWEALKYWWNLLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGAYRAIRHIPR
                     RIRQGLERILL"
BASE COUNT     3431 a   1781 c   2368 g   2168 t
ORIGIN

Initial Score    =    664  Optimized Score  =    671  Significance = 52.01
Residue Identity =    97%  Matches          =    673  Mismatches   =    13
Gaps             =      3  Conservative Substitutions               =     0

          10        20        30        40        50        60        70
    GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          X       10        20        30        40        50        60

       80        90       100       110       120       130       140
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
    |||||||||||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
    GGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
       70        80        90       100       110       120       130

      150       160       170       180       190       200       210
    TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
    |||||||||||||||||||||||||| ||| |||||||| ||||| ||||||||||||||||||||||||||
    TACAAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAAGCCAACAAAGGAGAGAACACCAGCTTGTTACAC
    140       150       160       170       180       190       200

     220       230       240       250       260       270       280
```

```
   CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
   |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
   CCTGTGAGCCTGCATGGAATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
 210       220       230       240       250       260       270       280

  290       300       310       320       330       340       350       360
   TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
   ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
         290       300       310       320       330       340       350

          370       380       390       400       410       420       430
   CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
   ||||||||||||| |||||||||||||||||||||||||||||| |||||||||||||||||||||||| ||||
   CTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGC
       360       370       380       390       400       410       420

        440       450       460       470       480       490       500
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
     430       440       450       460       470       480       490

      510       520       530       540       550       560       570
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   500       510       520       530       540       550       560

    580       590       600       610       620       630       640
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
 570       580       590       600       610       620       630       640

  650       660       670       680       690     X
   CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
   |||||||||||| ||||||||||||||||||||   ||||||||||||
   CCGAACAGGGACCTGAAAGCGAAAGGGAAACCA---GAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
         650       660       670          680     X 690       700       710

   CGCACGGCAAGAGGCGAGGGGCGGCG
          720       730
```

4. RAILEY-000-716.SE@ (1-696)
   HIVH3CG Human T-cell lymphotropic virus type III, complete

```
ID   HIVH3CG    standard; RNA; VRL; 9749 BP.
XX
AC   K02010; K02008; K02009;
XX
DT   18-NOV-1986 (Rel. 10, Created)
DT   23-OCT-1992 (Rel. 33, Last updated, Version 4)
XX
DE   Human T-cell lymphotropic virus type III, complete reference genome
DE   (isolates HXB2, HXB3, BH10, BH5 and BH8 of HTLV-III DNA).
XX
KW   acquired immune deficiency syndrome; complete genome; env gene;
KW   gag gene; long terminal repeat; pol gene; polyprotein; provirus;
KW   reverse transcriptase; tar protein; trans-activator.
XX
OS   Human immunodeficiency virus type 1
OC   Viridae; ss-RNA enveloped viruses; Positive strand RNA viruses;
OC   Retroviridae; Lentivirinae.
XX
```

RN [1]
RP 1-653, 9116-9749
RA Starcich B., Ratner L., Josephs S.F., Okamato T., Gallo R.C.,
RA Wong-staal F.;
RT "Characterization of long terminal repeat sequences of HTLV-III";
RL Science 227:538-540(1985).
XX
RN [2]
RP 1-9749
RA Wong-staal F., Gallo R.C., Chang N.T., Ghrayeb J., Papas T.S.,
RA Lautenberger J.A., Pearson M.L., Petteway S.R.Jr., Ivanoff L.,
RA Baumeister K., Whitehorn E.A., Rafalski J.A., Doran E.R.,
RA Josephs S.J., Starcich B., Livak K.J., Patarca R., Haseltine W.,
RA Ratner L.;
RT "Complete nucleotide sequence of the AIDS virus, HTLV-III";
RL Nature 313:277-284(1985).
XX
RN [3]
RC exons only, tat mrna
RP 508-9666
RA Arya S.K., Guo C., Josephs S.F., Wong-staal F.;
RT "Trans-activator gene of human T-lymphotropic virus type III
RT (HTLV-III)";
RL Science 229:69-73(1985).
XX
RN [4]
RP 5775-6082, 8397-8499
RA Sodroski J.G., Patarca R., Rosen C.A., Wong-staal F., Haseltine W.;
RT "Location of the trans-activating region on the genome of human
RT T-cell lymphotropic virus type III";
RL Science 229:74-77(1985).
XX
RN [5]
RC mrna splice sites
RA Rabson A.B., Daugherty D.F., Venkatesan S., Boulukos K.e.,
RA Benn S.I., Folks T.M., Feorino P., Martin M.;
RT "Transcription of novel open reading frames of AIDS retrovirus
RT during infection of lymphocytes";
RL Science 229:1388-1390(1985).
XX
RN [6]
RC 27k antigen cds
RA Allan J.S., Coligan J.E., Lee T.H., McLane M.F., Kanki P.J.,
RA Groopman J.E., Essex M.;
RT "A new HTLV-III/LAV encoded antigen detected by antibodies from
RT AIDS patients";
RL Science 230:810-813(1985).
XX
RN [7]
RC in hxb-3
RP 5778-8933
RA Crowl R., Ganguly K., Gordon M., Conroy R., Schaber M., Kramer R.,
RA Shaw G., Wong-staal F., Reddy E.P.;
RT "HTLV-III env gene products synthesized in E. coli are recognized
RT by antibodies present in the sera of AIDS patients";
RL Cell 41:979-986(1985).
XX
RN [8]
RC gp160 and gp120 coding sequences
RA Allan J.S., Coligan J.E., Barin F., McLane M.F., Sodroski J.G.,
RA Rosen C.A., Haseltine W.A., Lee T.H., Essex M.;
RT "Major glycoprotein antigens that induce antibodies in AIDS
RT patients are encoded by HTLV-III";
RL Science 228:1091-1094(1985).
XX
RN [9]

RC regulatory sequences in the ltr
RA Rosen C.A., Sodroski J.G., Haseltine W.A.;
RT "The location of cis-acting regulatory sequences in the human T
RT cell lymphotropic virus type III (HTLV-III/LAV) long terminal
RT repeat";
RL Cell 41:813-823(1985).
XX
RN [10]
RP 1-9749
RA Van Beveren C., Coffin J.M., Hughes S.;
RT "Appendix B: HTLV-3/LAV genome";
RL (in) Weiss R., Teich N., Varmus and Coffin J.M. (eds.);
RL RNA TUMOR VIRUSES SECOND EDITION:1102-1148;
RL Cold Spring Harbor Laboratory, New York (1985)
XX
RN [11]
RC trans-activator function and tar sequence
RA Rosen C.A., Sodroski J.G., Goh W.C., Dayton A.I., Lippke J.,
RA Haseltine W.A.;
RT "Post-transcriptional regulation accounts for the trans-activation
RT of the human T-lymphotropic virus type III";
RL Nature 319:555-559(1986).
XX
RN [12]
RC pol coding sequence
RA Marzo Veronese F., Copeland T.D., DeVico A.L., Rahman R.,
RA Oroszlan S., Gallo R.C., Sarngadharan M.G.;
RT "Characterization of highly immunogenic p66/p51 as the reverse
RT transcriptase of HTLV-III/LAV";
RL Science 231:1289-1291(1986).
XX
RN [13]
RC the 23k sor gene product
RA Kan N.C., Franchini G., Wong-staal F., DuBois G.C., Robey W.G.,
RA Lautenberger J.A., Papas T.S.;
RT "Identification of HTLV-III/LAV sor gene product and detection of
RT antibodies in human sera";
RL Science 231:1553-1555(1986).
XX
RN [14]
RC pol nh2-terminal region
RA Kramer R.A., Schaber M.D., Skalka A.M., Ganguly K., Wong-staal F.,
RA Reddy E.P.;
RT "HTLV-III gag protein is processed in yeast cells by the virus
RT pol-protease";
RL Science 231:1580-1584(1986).
XX
RN [15]
RC sor 23k protein
RA Lee T.H., Coligan J.E., Allan J.S., McLane M.F., Groopman J.E.,
RA Essex M.;
RT "A new HTLV-III/LAV protein encoded by a gene found in cytopathic
RT retroviruses";
RL Science 231:1546-1549(1986).
XX
RN [16]
RC sor 23k protein
RA Sodroski J.G., Goh W.C., Rosen C.A., Tartar A., Portetelle D.,
RA Burny A., Haseltine W.;
RT "Replicative and cytopathic potential of HTLV-III/LAV with sor
RT gene deletions";
RL Science 231:1549-1553(1986).
XX
RN [17]
RC sp1 binding sites in the promoter region
RA Jones K.A., Kadonaga J.T., Luciw P.A., Tjian R.;

RT "Activation of the AIDS retrovirus promoter by the cellular
RT transcription factor, Sp1";
RL Science 232:755-759(1986).
XX
RN [18]
RC acceptor and donor splice sites for tat and 27k
RA Arya S.K., Gallo R.C.;
RT "Three novel genes of human T-lymphotropic virus type III: Immune
RT reactivity of their products with sera from acquired immune
RT deficiency syndrome patients";
RL Proc. Natl. Acad. Sci. U.S.A. 83:2209-2213(1986).
XX
RN [19]
RC deletion mutants in the tat gene
RA Dayton A.I., Sodroski J.G., Rosen C.A., Goh W.C., Haseltine W.A.;
RT "The trans-activator gene of the human T cell lymphotropic virus
RT type III is required for replication";
RL Cell 44:941-947(1986).
XX
RN [20]
RC hypervariable and conserved regions in the env gene
RA Willey R.W., Ruthledge R.A., Dias S., Folks T., Theodore T.S.,
RA Buckler C.E., Martin M.A.;
RT "Identification of conserved and divergent domains within the
RT envelope gene of the acquired immunodeficiency syndrom
RT retrovirus";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5038-5042(1986).
XX
RN [21]
RC art cds boundaries
RA Sodroski J.G., Goh W.C., Rosen C.A., Dayton A., Terwilliger E.,
RA Haseltine W.;
RT "A second post-transcriptional trans-activator gene required for
RT HTLV-III replication";
RL Nature 321:412-417(1986).
XX
DR EPD; 14085; HIV-1(HTLV-III) LTR.
DR SWISS-PROT; P03347; GAG_HIV10.
DR SWISS-PROT; P03366; POL_HIV10.
DR SWISS-PROT; P03375; ENV_HIV10.
DR SWISS-PROT; P03401; VIF_HIV10.
DR SWISS-PROT; P03404; NEF_HIV10.
DR SWISS-PROT; P04606; TAT_HIV10.
DR SWISS-PROT; P04616; REV_HIV10.
DR SWISS-PROT; P04617; REV_HIV1P.
DR SWISS-PROT; P04624; ENV_HIV1Y.
DR SWISS-PROT; P05854; NEF_HIV1Y.
DR SWISS-PROT; P05920; VPU_HIV10.
DR SWISS-PROT; P05926; VPR_HIV10.
XX
CC Sequence for [7] was kindly supplied in computer readable form by
CC R. Crowl, 09/17/85. R. Patarca provided sites information and a
CC clean copy for [4], 09/16/85. Acquired immune deficiency syndrome
CC (AIDS) is caused by a retrovirus known by several names, perhaps
CC representing two separate strains: human T-cell lymphotropic
CC virus-III (HTLV-III), whose sequence is given below, and
CC lymphadenopathy-associated virus (LAV) are thought to be one strain
CC differing from AIDS-associated retrovirus type 2 (ARV-2) when
CC overall homology is the criterion. Some reading frame similarities
CC suggest that ARV-2 and LAV are more closely related. All three
CC viruses, whose sequences do not differ by more than 6%, are
CC believed to belong to the C type subfamily Lentiviridae, the "slow"
CC retroviruses. The BH10 sequence differs from BH8 and BH5 by 0.9% in
CC the coding regions and 1.8% in the noncoding regions, and the
CC authors of [2] believe that these are stable variants. The 5' and
CC 3' LTRs of BH10 and BH8 were not fully sequenced; the missing bases

CC (493-675 and 9608-9749) were filled in by [2] from the proviral
CC clone HXB2 [1]. The sequence below is that of BH10 with exception
CC of the variation at position 9197 which allows annotation of the
CC 27K coding sequence. The BH8 sequence spans bases 6033 to 9607, the
CC BH5 sequence spans bases 675 to 6038, and the HXB3 sequence [7]
CC spans bases 5778 to 8933. While this entry is offerred as the
CC reference locus for the AIDS retroviral sequence loci, no claim is
CC being made that this sequence is more prevalent or typical than
CC others, all of which have been entered in this library with
CC annotation. The HTLV-III genome encodes at least six proteins or
CC polyproteins: gag, pol, env, TAT, 27K antigen and the sor 23K
CC product. The 3' ORF (positions 8797-9447) is truncated in BH10
CC (stop codon at positions 9196-9198), but reads through in BH8 and
CC other sequences to yield what is now called the 27K antigen. The
CC sequence below is from BH10 with exception of the variation at
CC position 9197 which allows annotation of the 27K coding sequence.
CC Additionally there are four short open reading frames, bases
CC 1248-1406, 4442-4642, 5592-5828 and 6095-6340, which are conserved
CC to a large degree. A seventh gene has been proposed based upon a
CC combination of mutational and regulatory evidence: called "ART" (
CC for anti-repression transactivator), its product appears to act
CC post-transcriptionally to relieve negative repression of gag and
CC env production [21]. The exon assignments for ART are putative, but
CC if they are corroborated, the ART protein would be 116 amino acids
CC in length. The mechanism for pol gene translation has not been
CC elucidated: a gag-pol fusion protein is possible; splicing or
CC frameshift have not been ruled out. The viral protease would be
CC determined by the region in question. Approximately two-thirds of
CC the variant sites in the gag and pol genes are "silent mutations",
CC while over half of those in the env gene are not. Reference [20]
CC defines divergent and conserved regions for the env gene. Because
CC of the excessive variability of the env gene, differences between
CC the sequences summarized herein and other env gene entries have not
CC been annotated; only HTLV-III sequence variations have been
CC included in the sites of this entry. Other entries will include
CC information for alignment with this entry, including the Zaire and
CC New York isolate sequences reported by [20]. The TAT protein
CC (trans-activator protein, approximately 14 kd) is an effector of an
CC autostimulatory pathway through interaction with a positive control
CC element, the trans-activating responsive sequence, TAR. TAT seems
CC to be a transcriptional control molecule in HTLV-I, but [11]
CC demonstrates that it is a post-transcriptional regulatory molecule
CC in HTLV-III. Deletion mutants in the TAT gene are incapable of
CC prolific replication and exhibit no cytopathic effects in T4+ cell
CC lines [19]. The TAR sequence(s) are found to be between -17 and +80
CC relative to the cap site +1 (base 455) and is highly conserved.
CC Enhancer sequences which need not be viral-specific are found
CC upstream from TAR [9],[11]. Three tandem decanucleotide Sp1 binding
CC sites are located between bases 377 and 409, of which site III
CC shows the strongest affinity for the cellular factor; intact, the
CC three sites cause up to a tenfold effect on transcriptional
CC efficiency in vitro ([17] (The authors demonstrate the existence of
CC Sp1 in a human T-cell line). In addition to the ~9.4 kb genomic
CC mRNA, subgenomic mRNAs of 7.4, 5.5, 5.0, 4.3, 2.0 and 1.8 have been
CC detected. All are probably polyadenylated at the same site,
CC position 9666 below, with a potential polyadenyation signal at
CC 9642-9648, and capped at the same site, position 455, with a
CC potential TATA box at 427-431. The doubly-spliced transcript of
CC about 2.0 kb is responsible for the TAT message at least, and
CC depending upon the acceptor site, also for the sor and 27K
CC messages, given that a single, albeit partial, mRNA exists for all
CC three [18]. The acceptor splice for TAT is at position 5811 and the
CC putative acceptor splice for 27K is at position 6010; the donor
CC splice site in all three cases would be at position 6079 [18]. The
CC doubly spliced message would also encode the newly proposed ART
CC protein.

```
XX
FH   Key             Location/Qualifiers
FH
FT   repeat_region   1..634
FT                   /note="5' LTR"
FT   repeat_region   1..634
FT                   /note="5' LTR"
FT   variation       82..82
FT                   /note="a in BH10; g in H9"
FT   variation       101..101
FT                   /note="g in BH10; a in H9"
FT   variation       108..108
FT                   /note="a in [2], H9; g in HXB2 [1]"
FT   variation       164..164
FT                   /note="g in [2]; t in HXB2 [1], H9"
FT   variation       168..168
FT                   /note="t in [2]; g in HXB2 [1], H9"
FT   variation       176..176
FT                   /note="a in [2]; g in HXB2 [1], H9"
FT   variation       183..183
FT                   /note="c in [2], H9; t in HXB2 [1]"
FT   variation       227..227
FT                   /note="a in [2], H9; g in HXB2 [1]"
FT   variation       291..291
FT                   /note="a in [2]; g in HXB2 [1], H9"
FT   variation       333..333
FT                   /note="c in [2]; t in HXB2 [1], H9"
FT   misc_feature    377..386
FT                   /note="Sp1 binding site III [17]"
FT   misc_feature    388..397
FT                   /note="Sp1 binding site II [17]"
FT   misc_feature    399..408
FT                   /note="Sp1 binding site I [17]"
FT   variation       421..421
FT                   /note="c in BH10, BH5; t in H9"
FT   repeat_region   454..551
FT                   /note="R repeat 5' copy"
FT   repeat_region   454..551
FT                   /note="R repeat 5' copy"
FT   misc_RNA        455..455
FT                   /note="genomic mRNA start (cap site) [10]"
FT   misc_RNA        455..455
FT                   /note="TAT,ART mRNA exon 1 start (cap site) [10],
FT                   [18],[21]"
FT   variation       501..501
FT                   /note="a in BH10, BH5, H9; g in HXB2 [1]"
FT   misc_feature    636..653
FT                   /note="primer (Lys-tRNA) binding site"
FT   variation       654..654
FT                   /note="c in BH10, BH5; t in H9"
FT   variation       677..677
FT                   /note="g in BH10, BH5; ggag in H9"
FT   variation       704..704
FT                   /note="tga in BH10, H9; g in BH5 [2]"
FT   CDS             787..2325
FT                   /note="gag polyprotein precursor"
FT   variation       1290..1290
FT                   /note="a in BH10; g in BH5 [2], H9"
FT   variation       1431..1431
FT                   /note="a in BH10; g in BH5 [2], H9"
FT   variation       1455..1455
FT                   /note="t in BH10, H9; c in BH5 [2]"
FT   variation       1611..1611
FT                   /note="a in BH10, H9; g in BH5 [2]"
FT   variation       1620..1620
FT                   /note="c in BH10, H9; t in BH5 [2]"
```

```
FT   variation       1656..1656
FT                   /note="a in BH10, H9; g in BH5 [2]"
FT   variation       1662..1662
FT                   /note="t in BH10; c in BH5 [2], H9"
FT   variation       1675..1675
FT                   /note="g in BH10, BH5; c in H9"
FT   variation       1722..1722
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       1806..1806
FT                   /note="g in BH10, BH5; a in H9"
FT   variation       1845..1845
FT                   /note="a in BH10, BH5; g in H9"
FT   variation       1903..1903
FT                   /note="a in BH10, H9; t in BH5 [2]"
FT   variation       1906..1906
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       1923..1923
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       1950..1950
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       1953..1953
FT                   /note="g in BH10, H9; t in BH5 [2]"
FT   variation       1988..1988
FT                   /note="c in BH10, H9; t in BH5 [2]"
FT   variation       1992..1992
FT                   /note="c in BH10, H9; a in BH5 [2]"
FT   variation       2003..2003
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       2013..2013
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   CDS             2391..5129
FT                   /note="pol polyprotein (NH2-terminus uncertain; AA
FT                   at 2391)"
FT   variation       2468..2468
FT                   /note="g in BH10, BH5; a in H9"
FT   variation       2591..2591
FT                   /note="c in BH10, H9; t in BH5 [2]"
FT   variation       2600..2600
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       2741..2741
FT                   /note="g in BH10; a in BH5 [2], H9"
FT   variation       2827..2827
FT                   /note="a in BH10, H9; g in BH5 [2]"
FT   variation       2858..2858
FT                   /note="a in BH10, H9; g in BH5 [2]"
FT   variation       2990..2990
FT                   /note="c in BH10, H9; t in BH5 [2]"
FT   variation       3007..3007
FT                   /note="tta in BH10, H9; gtg in BH5 [2]"
FT   variation       3097..3097
FT                   /note="a in BH10; g in BH5 [2], H9"
FT   variation       3122..3122
FT                   /note="c in BH10, H9; t in BH5 [2]"
FT   variation       3222..3222
FT                   /note="c in BH10, H9; t in BH5 [2]"
FT   variation       3302..3302
FT                   /note="ag in BH10, H9; ga in BH5 [2]"
FT   variation       3368..3368
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       3389..3389
FT                   /note="g in BH10, BH5; a in H9"
FT   variation       3395..3395
FT                   /note="c in BH10, H9; t in BH5 [2]"
FT   variation       3755..3755
FT                   /note="a in BH10, BH5; g in H9"
FT   variation       3767..3767
```

```
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       3833..3833
FT                   /note="t in BH10, BH5; c in H9"
FT   variation       3855..3855
FT                   /note="t in BH10, BH5; c in H9"
FT   variation       3899..3899
FT                   /note="c in BH10, BH5; t in H9"
FT   variation       3922..3922
FT                   /note="a in BH10, H9; g in BH5 [2]"
FT   variation       3934..3934
FT                   /note="a in BH10, BH5; g in H9"
FT   variation       3954..3954
FT                   /note="g in BH10, BH5; c in H"
FT   variation       3962..3962
FT                   /note="caa in BH10, H9; tag in BH5 [2]"
FT   variation       3977..3977
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       3984..3984
FT                   /note="c in BH10, H9; a in BH5 [2]"
FT   variation       3993..3993
FT                   /note="a in BH10, H9; c in BH5 [2]"
FT   variation       4010..4010
FT                   /note="a in BH10; g in BH5 [2], H9"
FT   variation       4016..4016
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       4029..4029
FT                   /note="t in BH10, H9; c in BH5 [2]"
FT   variation       4049..4049
FT                   /note="a in BH10; g in BH5 [2], H9"
FT   variation       4064..4064
FT                   /note="c in BH10, H9; t in BH5 [2]"
FT   variation       4116..4116
FT                   /note="a in BH10, BH5; c in H9"
FT   variation       4167..4167
FT                   /note="g in BH10, BH5; c in H9"
FT   variation       4292..4292
FT                   /note="t in BH10, H9; a in BH5 [2]"
FT   CDS             5074..5652
FT                   /note="sor 23K protein"
FT   variation       5156..5156
FT                   /note="a in BH10, H9; g in BH5 [2]"
FT   variation       5314..5314
FT                   /note="t in BH10, BH5; c in H9"
FT   variation       5348..5348
FT                   /note="a in BH10, H9; g in BH5 [2]"
FT   variation       5401..5401
FT                   /note="t in BH10, H9; c in BH5 [2]"
FT   variation       5412..5412
FT                   /note="c in BH10, H9; t in BH5 [2]"
FT   variation       5548..5548
FT                   /note="a in BH10, H9; g in BH5 [2]"
FT   variation       5628..5628
FT                   /note="g in BH10, H9; a in BH5 [2]"
FT   variation       5846..5846
FT                   /note="g in BH10, H9, HXB3; a in BH5 [2]"
FT   CDS             5864..6078
FT                   /note="TAT protein,exon 2 (first expressed exon)"
FT   variation       5934..5934
FT                   /note="a in BH10, H9, HXB3; c in BH5 [2]"
FT   CDS             6003..6078
FT                   /note="ART protein,exon 2 (first expressed exon;
FT                   putative)"
FT   variation       6035..6045
FT                   /note="cctcctcaagg in BH10,HXB3 [7]; gctcatcgaag
FT                   in BH8 [2]; g in BH5 [2],clone 12 cDNA [21]"
FT   variation       6086..6086
```

```
FT                   /note="g in BH10, BH8, H9; a in HXB3 [7]"
FT   variation       6096..6096
FT                   /note="t in BH10, HXB3 [7], H9; c in BH8 [2]"
FT   variation       6108..6108
FT                   /note="a in BH10, HXB3 [7], H9; c in BH8 [2]"
FT   variation       6113..6114
FT                   /note="gc in BH10,HXB3 [7],H9; gtaac in BH8 [2]"
FT   variation       6124..6124
FT                   /note="a in BH10, HXB3 [7], H9; c in BH8 [2]"
FT   variation       6152..6152
FT                   /note="g in BH10, HXB3 [7], BH8; c in H9"
FT   CDS             6255..8825
FT                   /note="envelope protein precursor (env)"
FT   variation       6373..6373
FT                   /note="a in BH10, HXB3 [7], H9; t in BH8 [2]"
FT   variation       6474..6474
FT                   /note="t in BH10, BH8 [2], H9; g in HXB3 [7]"
FT   variation       6748..6748
FT                   /note="t in BH10, HXB3 [7], H9; a in BH8 [2]"
FT   variation       6929..6929
FT                   /note="t in BH10, HXB3 [7], H9; c in BH8 [2]"
FT   variation       7088..7088
FT                   /note="a in BH10, H9; g in BH8 [2], HXB3 [7]"
FT   variation       7119..7119
FT                   /note="a in BH10; HXB3 [7], H9; g in BH8 [2]"
FT   variation       7121..7123
FT                   /note="cca in BH10,H9; cac in BH8 [2],HXB3 [7]"
FT   variation       7171..7172
FT                   /note="gt in BH10, H9; aa in BH8 [2], HXB3[7]"
FT   variation       7187..7187
FT                   /note="a in BH10, H9; g in BH8 [2], HXB3 [7]"
FT   variation       7272..7273
FT                   /note="aa in BH10, H9; gc in BH8[2], HXB3 [7]"
FT   variation       7291..7291
FT                   /note="a in BH10, BH8 [2], H9; c in HXB3 [7]"
FT   variation       7343..7343
FT                   /note="g in BH10, BH8 [2]; a in HXB3 [7], H9"
FT   variation       7439..7454
FT                   /note="gtttaatagtacttgg in BH10,HXB3 [7],and H9"
FT   variation       7461..7461
FT                   /note="a in BH10, BH8 [2]; g in HXB3 [7], H9"
FT   variation       7499..7499
FT                   /note="c in BH10, BH8 [2]; a in HXB3 [7], H9"
FT   variation       7521..7521
FT                   /note="a in BH10, BH8 [2]; t in HXB3 [7], H9"
FT   variation       7574..7574
FT                   /note="t in BH10, CH8 [2]; c in HXB3 [7], H9"
FT   variation       7636..7637
FT                   /note="cg in BH10, HXB3 [7], H9; gc in BH8[2]"
FT   variation       7636..7636
FT                   /note="g in BH10, BH8 [2]; a in HXB3 [7], H9"
FT   variation       7645..7645
FT                   /note="a in BH10, BH8 [2], H9; g in HXB3 [7]"
FT   variation       8060..8061
FT                   /note="ca in BH10, BH8 [2], H9; ac in H"
FT   variation       8127..8127
FT                   /note="a in BH10, BH8 [2], H9; c in HXB[7]"
FT   variation       8131..8131
FT                   /note="t in BH10, BH8 [2], H9; c in HXB3 [7]"
FT   variation       8135..8135
FT                   /note="c in BH10, BH8 [2], H9; g in HXB3 [7]"
FT   variation       8257..8257
FT                   /note="g in BH10, BH8, HXB3; a in H9"
FT   variation       8273..8273
FT                   /note="t in BH10, BH8, HXB3; g in H9"
FT   variation       8364..8364
```

```
FT                   /note="g in BH10, HXB3 [7]; a in BH8 [2], H9"
FT   CDS             8409..8683
FT                   /note="ART protein,exon 3 (putative; AA at 8411)"
FT   CDS             8409..8454
FT                   /note="TAT protein, exon 3 ( AA at 8410)"
FT   variation       8422..8422
FT                   /note="t in BH10,HXB3 [7],clone 12 cDNA [21]; a in
FT                   BH8 [2]; c in H9"
FT   variation       8464..8464
FT                   /note="g in BH10,BH8,HXB3,clone 12 cDNA [21]; a in
FT                   H9"
FT   variation       8657..8657
FT                   /note="g in BH10,BH8 [2]; a in HXB3 [7],H9,clone
FT                   12 cDNA [21]"
FT   variation       8672..8672
FT                   /note="g in BH10,HXB3 [7],clone 12 cDNA [21],H9; a
FT                   in BH8 [2]"
FT   variation       8692..8692
FT                   /note="g in BH10,HXB3 [7],clone 12 cDNA [21],H9; a
FT                   in BH8 [2]"
FT   variation       8748..8748
FT                   /note="g in BH10,HXB3 [7],clone 12 cDNA [21],H9; t
FT                   in BH8 [2]"
FT   variation       8758..8758
FT                   /note="g in BH10,H9; c in BH8 [2]; a in HXB3 [7],
FT                   clone 12 cDNA [21]"
FT   variation       8771..8771
FT                   /note="t in BH10,HXB3 [7],clone 12 cDNA [21],H9; c
FT                   in BH8 [2]"
FT   CDS             8827..9447
FT                   /note="27K protein,exon 3 (first expressed exon)"
FT   variation       8857..8857
FT                   /note="g in BH10,BH8,HXB3,clone 12 cDNA [21]; a in
FT                   H9"
FT   variation       8924..8924
FT                   /note="c in BH10,HXB3 [7],clone 12 cDNA [21],H9; t
FT                   in BH8 [2]"
FT   variation       8967..8967
FT                   /note="c in BH10,clone 12 cDNA [21],H9; t in BH8
FT                   [2]"
FT   variation       8978..8978
FT                   /note="a in BH10,clone 12 cDNA [21],H9; c in BH8
FT                   [2]"
FT   variation       8985..8985
FT                   /note="t in BH10,clone 12 cDNA [21],H9; c in BH8
FT                   [2]"
FT   variation       8987..8987
FT                   /note="a in BH10,BH8; c in H9,clone 12 cDNA [21]"
FT   variation       8994..8994
FT                   /note="c in BH10,clone 12 cDNA [21],H9; t in BH8
FT                   [2]"
FT   variation       9019..9019
FT                   /note="g in BH10,BH8; a in H9,clone 12 cDNA [21]"
FT   repeat_region   9116..9749
FT                   /note="3' LTR"
FT   variation       9169..9196
FT                   /note="t in BH10,clone 12 cDNA [21]; c in BH8 [2]"
FT   variation       9197..9197
FT                   /note="g in BH8 [2],H9,clone 12 cDNA [21]; a in
FT                   BH10 [2]"
FT   variation       9216..9216
FT                   /note="g in BH10,BH8; a in H9,clone 12 cDNA [21]"
FT   variation       9222..9223
FT                   /note="ga in BH10,clone 12 cDNA [21],H9; ag in
FT                   BH8[2]"
FT   variation       9279..9279
```

```
FT                   /note="g in BH10,BH8,clone 12 cDNA [21]; t in H9"
FT   variation       9283..9283
FT                   /note="t in BH10,BH8,clone 12 cDNA [21]; g in H9"
FT   variation       9284..9284
FT                   /note="t in BH10,H9,clone 12 cDNA [21]; a in BH8
FT                   [2]"
FT   variation       9291..9291
FT                   /note="a in BH10,BH8,clone 12 cDNA [21]; g in H9"
FT   variation       9297..9297
FT                   /note="c in BH10,clone 12 cDNA [21],H9; t in BH8
FT                   [2]"
FT   variation       9354..9354
FT                   /note="g in BH10, HIVDSM>, H9; t in BH8 [2]"
FT   variation       9406..9406
FT                   /note="a in BH10,BH8; g in H9,clone 12 cDNA [21]"
FT   variation       9448..9448
FT                   /note="c in BH10; t in BH8 [2],H9,clone 12 cDNA"
FT   variation       9536..9563
FT                   /note="c in BH10,BH8,clone 12 cDNA [21]; g in H9"
FT   repeat_region   9570..9666
FT                   /note="R repeat 3' copy"
FT   variation       9616..9616
FT                   /note="g in HXB2; a in H9, clone 12 cDNA [21]"
FT   variation       9621..9621
FT                   /note="g in HXB2; a in H9, clone 12 cDNA [21]"
FT   variation       9663..9663
FT                   /note="t in BH10,H9; tg in clone 12 cDNA [21]"
FT   polyA_site      9666..9666
FT                   /note="TAT,ART,27K mRNA exon 3 end (poly-A site)
FT                   [10],[18],[21]"
FT   polyA_site      9666..9666
FT                   /note="genomic mRNA end (poly-A site) [10]"
XX
SQ   Sequence 9749 BP; 3431 A; 1781 C; 2369 G; 2168 T; 0 other;

Initial Score    =    664  Optimized Score  =    671  Significance = 52.01
Residue Identity =    97%  Matches          =    673  Mismatches   =    13
Gaps             =      3  Conservative Substitutions              =     0

        10        20        30        40        50        60        70
    GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          X       10        20        30        40        50        60

         80        90       100       110       120       130       140
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
    |||||||||||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
    GGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
       70        80        90       100       110       120       130

       150       160       170       180       190       200       210
    TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
    |||||||||||||||||||||||||| ||| ||||||| ||||| ||||||||||||||||||||||||||
    TACAAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAAGCCAACAAAGGAGAGAACACCAGCTTGTTACAC
    140       150       160       170       180       190       200

     220       230       240       250       260       270       280
    CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
    |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
    CCTGTGAGCCTGCATGGAATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
  210       220       230       240       250       260       270       280

   290       300       310       320       330       340       350       360
    TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
    ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
   TTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
         290       300       310       320       330       340       350

          370       380       390       400       410       420       430
   CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
   |||||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||| ||||
   CTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGC
       360       370       380       390       400       410       420

         440       450       460       470       480       490       500
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
   |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
     430       440       450       460       470       480       490

        510       520       530       540       550       560       570
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   500       510       520       530       540       550       560

     580       590       600       610       620       630       640
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
 570       580       590       600       610       620       630       640

   650       660       670       680       690     X
    CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
    |||||||||||| ||||||||||||||||||||   ||||||||||||
    CCGAACAGGGACCTGAAAGCGAAAGGGAAACCA---GAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
          650       660       670          680     X 690       700       710

    CGCACGGCAAGAGGCGAGGGGCGGCG
           720       730
```

5. RAILEY-000-716.SEQ (1-696)
   HIVJRCSF Human immunodeficiency virus type 1, isolate JRCSF

```
LOCUS       HIVJRCSF     9540 bp ss-RNA             VRL       28-SEP-1992
DEFINITION  Human immunodeficiency virus type 1, isolate JRCSF; complete
            genome.
ACCESSION   M38429
KEYWORDS    long terminal repeat (LTR).
SOURCE      HIV-1 proviral DNA from extracellular virus taken from cerebral
            spinal fluid (1986). Infectious clone.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9540)
  AUTHORS   Koyanagi,S. and Chen,I.S.
  JOURNAL   Unpublished (1988) UCLA School of Medicine, Los Angeles.
  STANDARD  full automatic
COMMENT     Kindly provided in computer-readable form by Irvin Chen, UCLA
            School of Medicine, Los Angeles. JRCSF and JRFL (see <HIVJRFL> were
            isolated from cerebral spinal fluid and brain tissue of the patient
            JR, who died with Kaposi's sarcoma and severe AIDS encepha- lopathy
            (Science 236, 819-822, 1987). Both clones are infectious, but JRFL
            productively infects macrophages while JRCSF does not.  (Peripheral
            blood was not available from the patient).
            The JRCSF and JRFL env nucleotide sequences differ by at least 3%;
            further characterization of them is forthcoming (Peng,S. et al.,
            Nature 1990, in press). Both manifest insertions in nef previously
            reported for HIVBRVA.
FEATURES             Location/Qualifiers
```

```
LTR             1..635
                /partial
protein_bind    378..387
                /bound_moiety="Sp1"
protein_bind    389..398
                /bound_moiety="Sp1"
protein_bind    400..409
                /bound_moiety="Sp1"
exon            5842..6056
                /number=2
                /gene="tat"
exon            5981..6056
                /number=2
                /gene="rev"
exon            8366..8456
                /number=3
                /gene="tat"
exon            8366..8640
                /number=3
                /gene="rev"
LTR             9103..9540
                /partial
CDS             <2085..5108
                /gene="pol"
                /product="pol polyprotein"
                /codon_start=1
                /translation="FFREDLAFL@GKAREFPSE@TRANSPTRREL@VWGRDSNSLSEA
                GAEAGADR@GIVSFNFP@ITLW@RPLVTIKIGG@LKEALLDTGADDTVLEDMDLPGRW
                KPKMIGGIGGFIKVR@YD@IPIDICGHKAVGTVLVGPTPVNIIGRNLLT@IGCTLNFP
                ISPIETVPVKLKPGMDGPKVK@WPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTP
                VFAIKKKDSTKWRKLVDFRELNRRT@DFWEV@LGIPHPAGLKKKKSVTVLDVGDAYFS
                VPLDKDFRKYTAFTIPSINNETPGIRY@YNVLP@GWKGSPAIF@SSMTKILEPFRK@N
                PDIIIY@YMDDLYVGSDLEIG@HRTKIEELR@HLLKWGFTTPDKKH@KEPPFLWMGYE
                LHPDKWTV@PIVLPEKDSWTVNDI@KLVGKLNWAS@IYAGIKVK@LCKLLRGTKALTE
                VIPLTKEAELELAENREILKEPVHGVYYDPSKDLIVEI@K@GQG@WTY@IF@EPFKNL
                KTGKYARTRGAHTNDVK@LTEAV@KIANESIVIWGKIPKFKLPI@KETWETWWTEYWQ
                ATWIPEWEFVNTPPLVKLWY@LEKEPIVGAETFYVDGAANRETKLGKAGYVTSRGR@K
                VVSLTDTTN@KTEL@AIHLAL@DSGLEVNIVTDS@YALGII@A@PDKSESELVS@IIE
                @LIKKEKVYLAWVPAHKGIGGNE@VDKLVSAGIRKVLFLDGIDKA@EDHEKYHSNWRA
                MASDFNLPPIVAKEIVASCDKC@LKGEAMHG@VDCSPGIW@LDCTHLEGKIILVAVHV
                ASGYIEAEVIPAETG@ETAYFLLKLAGRWPVTTIHTDNGSNFTSTTVKAACWWAGIK@
                EFGIPYNP@S@GVVESMNKELKKIIG@VRD@AEHLKTAV@MAVFIHNFKRKGGIGGYS
                AGERIIDIIATDI@TKEL@K@ITKI@NFRVYYRDNRDPIWKGPAKLLWKGEGAVVI@D
                NSDIKVVPRRKVKIIRDYGK@MAGDDCVASR@DED"
CDS             790..2304
                /gene="gag"
                /product="gag polyprotein"
                /codon_start=1
                /translation="MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHIVWASRELERF
                AVNPGLLESSEGCR@ILG@L@PSLKTGSEELTSLYNTVATLYCVH@RIEIKDTKEALE
                KIEEE@TKSMKKAQ@AAADTGNSS@VS@NYPIV@NL@G@MVH@AISPRTLNAWVKVIE
                EKAFSPEVIPMFSALSEGATP@DLNTMLNTVGGH@AAM@MLKETINEEAAEWDRLHPV
                HAGPIAPG@MREPRGSDIAGTTSTL@E@IGWMTNNPPIPVGEIYKRWIILGLNKIVRM
                YSPVSILDIR@GPKEPFRDYVDRFYKTLRAE@AT@EVKNWMTETLLV@NANPDCKTIL
                KALGPAATLEEMMTAC@GVGGPGHKARVLAEAMS@VTNPATIMM@RGNFRN@RKNVKC
                FNCGKEGHIARNCRAPRKKGCWKCGKEGH@MKECTER@ANFLGKIWPSYKGRPGNFL@
                SRPEPTAPPEESFRFGEETATPS@K@E@K@EPIDKELYPLTSLRSLFGNDPSS@"
CDS             5053..5631
                /gene="vif"
                /product="vif protein"
                /codon_start=1
                /translation="MENRW@VMIVWQVDRMRIRTWNSLVKHHMYISGKAKGWIYKHHY
                ESTNPRVSSEV@IPLGDARLVITTYWGLHTGERDWHLG@GVSMEWRTRRYST@VDPDL
                AD@LIHLYYFDCFSESAIRNAILGHIVSPRCEY@AGHSKVGSL@YLALTALIKPKKIK
                PPLPSVKKLTEDRWNKPQKTKGHRGSHTMNGH"
CDS             5571..5861
```

```
                     /gene="vpR"
                     /codon_start=1
                     /translation="HEQAPEDQGPQREPYNEWTLELLEELKNEAVRHFPRIWLHSLGQ
                     YIYETYGDTWAGVEAIIRILQQLLFIHFRIGCRHSRIGITRQRRARNGASRS"
     CDS             6073..6318
                     /gene="vpU"
                     /codon_start=1
                     /translation="HQPLQILAIVALVVAGIIAIIVWSIVLIEYRKILRQRKIDRLID
                     KIRERAEDSGNESEGDQEELSALVERGHLAPWDINDL"
     CDS             6236..8782
                     /gene="env"
                     /product="envelope polyprotein"
                     /codon_start=1
                     /translation="MRVKGIRKNYQHLWKGGILLLGTLMICSAVEKLWVTVYYGVPVW
                     KETTTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLENVTEDFNMWKNNMVEQ
                     MQEDVINLWDQSLKPCVKLTPLCVTLNCKDVNATNTTSSSEGMMERGEIKNCSFNITK
                     SIRDKVQKEYALFYKLDVVPIDNKNNTKYRLISCNTSVITQACPKVSFEPIPIHYCAP
                     AGFAILKCNNKTFNGKGQCKNVSTVQCTHGIRPVVSTQLLLNGSLAEEKVVIRSDNFT
                     DNAKTIIVQLNESVKINCTRPSNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNISRAQ
                     WNNTLKQIVEKLREQFNNKTIVFTHSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWN
                     DTEKSSGTEGNDTIILPCRIKQIINMWQEVGKAMYAPPIKGQIRCSSNITGLLLTRDG
                     GKNESEIEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGIG
                     ALFLGFLGAAGSTMGARSMTLTVQARQLLSGIVQQQNNLLRAIEAQQHMLQLTVWGIK
                     QLQARVLAVERYLKDQQLMGIWGCSGKLICTTAVPWNTSWSNKSLDSIWNNMTWMEWE
                     KEIENYTNTIYTLIEESQIQQEKNEQELLELDKWASLWNWFGITKWLWYIKIFIMIVG
                     GLIGLRIVFSVLSIVNRVRQGYSPLSFQTLLPATRGPDRPEGIEEEGGERDRDRSGQL
                     VNGFLALIWVDLRSLFLFSYHRLRDLLLTVTRIVELLGRRGWEILKYWWNLLQYWSQE
                     LKNSAVSLLNATAIAVAEGTDRIIEVVQRVYRAILHIPTRIRQGLERALL"
     CDS             8784..9434
                     /gene="nef"
                     /codon_start=1
                     /translation="MGGKWSKHSVPGWSTVRERMRRAEPATDRVRQTEPAAVGVGAVS
                     RDLEKHGAITSSNTAATNADCAWLEAYEDEEVGFPVRPQVPLRPMTYKAAIDLSHFLK
                     EKGGLEGLIYSQKRQDILDLWIYHTQGYFPDWQNYTAGPGVRFPLTFGWCFKLVPVDP
                     EKVEEANEGENNCLLHPMSQHGMDDPEKEVLVWKFDSKLALHHVARELHPEYYKDC"
 BASE COUNT     3425 a   1691 c   2308 g   2116 t
 ORIGIN      5' terminus of 5'LTR.

Initial Score    =    652  Optimized Score  =    652  Significance = 51.03
Residue Identity =    94%  Matches          =    652  Mismatches   =    38
Gaps             =      0  Conservative Substitutions               =     0

         X 10        20        30        40        50        60        70
    GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          ||||||||||||||| |||| ||   ||||||||||||||||||||||||||||||||||||||||
          CTGGAAGGGCTAATTTACTCACAGAAAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          X       10        20        30        40        50        60

          80        90       100       110       120       130       140
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
    ||||||||||||||||||||||||||||||| |||| |||||||||||||| ||||||||||||||||||||||
    GGCTACTTCCCTGATTGGCAGAACTACACAGCAGGACCAGGGGTCAGATTTCCACTGACCTTTGGATGGTGC
      70        80        90       100       110       120       130

       150       160       170       180       190       200       210
    TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
    | |||||||||||||||||| ||||| |||||||||||||||||| |||||||||||| | ||||||||||
    TTCAAGCTAGTACCAGTTGATCCAGAGAAGGTAGAAGAGGCCAATGAAGGAGAGAACAACTGCTTGTTACAC
   140       150       160       170       180       190       200       210

     220       230       240       250       260       270       280
    CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
    ||| |||||| |||||||||||| ||||| |||| |||||||||| ||||| ||||||||||   ||||||
    CCTATGAGCCAGCATGGAATGGACGACCCAGAGAAGGAAGTGTTAGTGTGGAAGTTTGACAGCAAGCTAGCA
          220       230       240       250       260       270       280
```

```
  290       300       310       320       330       340       350       360
   TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
   || |||||||||||||||||||||||||||||||||| |||| |||||||||| ||||||| ||||||||||
   TTGCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTACAAGGACTGCTGACACCGAGCTTTCTACAAGGGA
        290       300       310       320       330       340       350

          370       380       390       400       410       420       430
   CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
   |||||||||||| |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
   CTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTGC
      360       370       380       390       400       410       420

        440       450       460       470       480       490       500
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
   |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
    430       440       450       460       470       480       490

      510       520       530       540       550       560       570
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CTAGCTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
  500       510       520       530       540       550       560       570

    580       590       600       610       620       630       640
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
         580       590       600       610       620       630       640

  650       660       670       680       690     X
   CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
   |||||||||||||  |||||||||||| |||||||||||||| ||||||||
   CCGAACAGGGACCGGAAAGCGAAAGAGAAACCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
        650       660       670       680       690       700       710

   CGCACAGCAAGAGGCGAGGGGCGGCG
      720       730       740
```

6. RAILEY-000-716.SEQ (1-696)
   HIVNY5CG Human immunodeficiency virus type 1, isolate NY5,

```
LOCUS       HIVNY5CG     9022 bp ss-RNA             VRL       28-SEP-1992
DEFINITION  Human immunodeficiency virus type 1, isolate NY5, complete genome.
            Infectious single LTR molecular proviral genome.
ACCESSION   M38431
KEYWORDS    .
SOURCE      HIV-1, isolate NY5, unintegrated circular viral DNA. Infectious.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9022)
  AUTHORS   Theodore,T. and Buckler-White,A.
  JOURNAL   Unpublished (1988)
  STANDARD  full automatic
COMMENT     Computer-readable copy of sequence kindly provided by Chuck
            Buckler, 01-NOV-1988. A partial sequence for NY5, isolated in 1984,
            is on page I-A-101 of this compendium and, as the 5' half of the
            hybrid HIVNL43, also an infectious clone, on page I-A-64.
            Hirt Supernatant DNA extracted from A3.01 cells infected with the
            NY5 HIV isolate stock was digested with EcoRI and cloned into
            lambda WESB. The insert is an EcoRI permuted single LTR clone and
            was then transferred into pBR322. In the sequence below position
            one is the first base of the single LTR of the clone while the last
            base (9022) is the one just before the LTR of the intact circle
```

```
FEATURES             Location/Qualifiers
     LTR             1..634
                     /partial
     protein_bind    377..386
                     /bound_moiety="Sp1"
     protein_bind    388..397
                     /bound_moiety="Sp1"
     protein_bind    399..408
                     /bound_moiety="Sp1"
     protein_bind    636..653
                     /bound_moiety="Lys-tRNA"
     exon            5830..6044
                     /number=2
                     /gene="tat"
     exon            5969..6044
                     /number=2
                     /gene="rev"
     exon            8316..8406
                     /number=3
                     /gene="tat"
     exon            8316..8590
                     /number=3
                     /gene="rev"
     CDS             <2085..5096
                     /gene="pol"
                     /codon_start=1
                     /translation="FFREDLAFPQGKAREFSSEQTRANSPTRREL0VWGRDNNSLSEA
                     GADRQGTVSFSFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMNLPGRWKPKM
                     IGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPI
                     ETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAI
                     KKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKQKKSVTVLDVGDAYFSVPLD
                     KDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQCSMTKILEPFRKQNPDIV
                     IYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPD
                     KWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYAGIKVRQLCKLLRGIKALTEVVPL
                     TEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGK
                     YARMKGAHTNDVKQLTEAVQKIATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWI
                     PEWEFVNTPPLVKLWYQLEKEPIIGAETFYVDGAANRETKLGKAGYVTDRGRQKVVPL
                     TDTTNQKTELQAIHLALQDSGLEVNIVTDSQYALGIIQAQPDKSESELVSQIIEQLIK
                     KEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVLFLDGIDKAQEEHEKYHSNWRAMASD
                     FNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGY
                     IEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSNFTSTTVKAACWWAGIKQEFGI
                     PYNPQSQGVIESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGER
                     IVDIIATDIQIKELQKQITKIQNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQDNSDI
                     KVVPRRKAKIIRDYGKQMAGDDCVASRQDED"
     CDS             790..2292
                     /gene="gag"
                     /codon_start=1
                     /translation="MGARASVLSGGELDKWEKIRLRPGGKKQYRLKHIVWASRELERF
                     AVNPGLLETSEGCRQILRQLQPSLQTGSEERRSLFNTVAVLYCVHQRIDVKDTKEALD
                     KIEEEQNKSKKKAQQAAADTGNSSQVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVVE
                     EKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPV
                     HAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRM
                     YSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTIL
                     KALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNPATIMIQRGNFRNQKKTVKC
                     FNCGKEGHIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQ
                     SRPEPTAPPEESFRFGEETTTPSQKQEPIDKELYPLASLRSLFGSDPSSQ"
     CDS             5041..5619
                     /gene="vif"
                     /codon_start=1
                     /translation="MENRWQVMIVWQVDRMRINTWKRLVKHHMYISRKAKDWFYRHHY
                     ESTNPKISSEVHIPLGDAKLVITTYWGLHTGERDWHLGQGVSIEWRKKRYSTQVDPDL
                     ADQLIHLHYFDCFSESAIRNTILGRIVSPRCEYQAGHNKVGSLQYLALAALIKPKQIK
                     PPLPSVRKLTEDRWNKPQKTKGHRGSHTMNGH"
     CDS             5559..5849
                     /gene="vpR"
                     /codon_start=1
```

```
                     /translation="MEQAPEDQGPQREPYNEWTLELLEELKSEAVRHFPRIWLHNLGQ
                     HIYETYGDTWAGVEAIIRILQQLLFIHFRIGCQHSRIGIIRQRRARNGASRS"
     CDS             6061..6147
                     /gene="vpU"
                     /codon_start=1
                     /translation="MLSFNSCVDHSVHRIQESIKTKKSRQDN"
     CDS             6180..8732
                     /gene="env"
                     /product="envelope polyprotein"
                     /codon_start=1
                     /translation="MRAKGTRKNYQHLWRWGTMLLGMLMICSAAEQLWVTVYYGVPVW
                     KEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLQNVTENFNMWKNNMVEQ
                     MHEDIISLWDQSLKPCVKLTPLCVTLNCTDLTNATYANGSSEERGEIRNCSFNVTTII
                     RNKIQKEYALFYRLDIVPIDKDNTSYTLINCDTSVITQACPKVSFEPIPIHYCAPAGF
                     AILKCNDKKFNGTGPCTNVSTVQCTHGIKPVVSTQLLLNGSLAEGEVVIRSENFTNNV
                     KTIIVQLNESVEINCTRPNNNTRKGIAIGPGRTLYARETIIGDIRQAHCNLSRAKWND
                     TLKQIVTKLKEQFRNKTIVFNQSSGGDPEIVMHSFNCGGEFFYCKTTQLFNSTWLFNS
                     TWNDTERSDNNETIILPCRIKQIINRWQEVGKAMYAPPISGQIRCSSNITGLLLTRDG
                     GDKENSTTEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGA
                     LGALFLGFLGAAGSTMGAASMALTVQTRQLMSGIVQQQNNLLKAIEAQQHLLQLTVWG
                     IKQLQARVLAVERYLKDQQLLRIWGCSGKLICTTTVPWNASWSNKSLDKIWDNMTWME
                     WEREIDNYTGLIYTLIEESQIQQEKNEQELLELDKWASLWNWFDITKWLWYIKIFIMI
                     VGGLIGLRIVFTVLSIVNRVRQGYSPLSFQTRLPAQRGPDRPEGIEEEGGERDRDRSG
                     PLVNGFLALIWVDLRSLCLFSYHRLRDLLLIVARIVELLGRRGWEALKYCWNLLQYWG
                     QELKNSAISLLNATAVAVAEGTDRVIEVAQRICRGILHIPRRIRQGLERLLL"
     CDS             8734..9022
                     /gene="nef"
                     /partial
                     /codon_start=1
                     /translation="MGGKWSKRSMSGWPAVRERMKRTEPAADGVGAVSRDLEKHGAIT
                     SSNTAATNANCAWLEAHEEEEVGFPVRPQVPLRPITRKAAMDLSHFLKEEGGX"
 BASE COUNT     3230 a   1591 c   2188 g   2013 t
 ORIGIN      5'terminus of 5'LTR (start of U3)

Initial Score    =    650  Optimized Score  =    650  Significance = 50.86
Residue Identity =    94%  Matches          =    650  Mismatches   =    39
Gaps             =      0  Conservative Substitutions               =     0

          10        20        30        40        50        60        70
    GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          ||||||||||||||   |||||| ||||||||||||||||||||||||||||||||||||||||||
          TGGAAGGGCTAATTTGGTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          X       10        20        30        40        50        60

        80        90       100       110       120       130       140
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
    ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
       70        80        90       100       110       120       130

       150       160       170       180       190       200       210
    TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
    | |||| ||||||||||||||||||   |||||||||||||||| |||||||||||| ||||||||||||||
    TTCAAGTTAGTACCAGTTGAGCCAGGGCAGGTAGAAGAGGCCAATGAAGGAGAGAACAACAGCTTGTTACAC
    140       150       160       170       180       190       200

     220       230       240       250       260       270       280
    CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
    ||| |||||| |||||| ||||| ||||| ||| |||||||| |||| ||||| |||||||||||| |||||||
    CCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGCA
  210       220       230       240       250       260       270       280

  290       300       310       320       330       340       350       360
    TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
    |||| |||| |||||||||||||||||||||||||| ||| |||||||||||||||||| ||||||||||
    TTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGA
```

```
        290       300       310       320       330       340       350

         370       380       390       400       410       420       430
    CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
    |||||||||||| ||||||||||||||| |||||||||||||| ||||||||||||||||||||||||||||||
    CTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTGC
        360       370       380       390       400       410       420

         440       450       460       470       480       490       500
    ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
    |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
    ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
      430       440       450       460       470       480       490

       510       520       530       540       550       560       570
    CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
    ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
    CTAGCTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTACAAGTAGTGTGTGCCCGTCT
    500       510       520       530       540       550       560

     580       590       600       610       620       630       640
    GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
  570       580       590       600       610       620       630       640

   650       660       670       680       690     X
    CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
    |||||||||||||||| |||||||| || ||||||||| ||||||||
    CCGAACAGGGACTTGAGAGCGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
          650       660       670       680       690       700       710

    CGCACGGCAAGAGGCGAGGGGCGGCG
        720       730
```

7. RAILEY-000-716.SEQ (1-696)
   HIVNL43 Human immunodeficiency virus type 1, NY5/BRU (LAV-

```
LOCUS       HIVNL43      9709 bp ss-RNA            VRL       15-JUN-1989
DEFINITION  Human immunodeficiency virus type 1, NY5/BRU (LAV-1) recombinant
            clone pNL4-3.
ACCESSION   M19921
KEYWORDS    .
SOURCE      Human immunodeficiency virus type 1 (HIV-1), NY5/BRU (LAV-1)
            recombinant clone pNL4-3.
  ORGANISM  Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Positive strand RNA virus;
            Retroviridae; Lentivirinae.
REFERENCE   1  (bases 1 to 9709)
  AUTHORS   Adachi,A., Gendelman,H.E., Koenig,S., Folks,T., Willey,R.,
            Rabson,A. and Martin,M.A.
  TITLE     Production of acquired immunodeficiency syndrome-associated
            retrovirus in human and nonhuman cells transfected with an
            infectious molecular clone
  JOURNAL   J. Virol. 59, 284-291 (1986)
  STANDARD  full automatic
REFERENCE   2  (bases 1 to 9709)
  AUTHORS   Buckler,C.E., Buckler-White,A.J., Willey,R.L. and McCoy,J.
  JOURNAL   Unpublished (1988) .
  STANDARD  full automatic
REFERENCE   3  (sites)
  AUTHORS   Buckler,C.E.
  JOURNAL   Unpublished (1988)
  STANDARD  full automatic
REFERENCE   4  (sites)
```

```
  AUTHORS   Dai,L.C., Littaua,R., Takahashi,K. and Ennis,F.A.
  TITLE     Mutation of human immunodeficiency virus type 1 at amino acid 585
            on gp41 resultis in loss of killing by CD8+ A24-restricted
            cytotoxic T lymphocytes
  JOURNAL   J. Virol. 66, 3151-3154 (1992)
  STANDARD  full automatic
COMMENT     [3] sites; revisions of [3].

            Clean copy of sequence [3] kindly provided by Chuck Buckler, NIAID,
            Bethesda, MD, 24-JUN-1988.  The construction of pNL4-3 has been
            described in [1].  pNL4-3 is a recombinant (infectious) proviral
            clone that contains DNA from HIV isolates NY5 (5' half) and BRU (3'
            half).  The site of recombination is the EcoRI site at positions
            5743-5748.

            The length and sequence of the vpr coding region corresponds to
            that of the BRU, SC, SF2, MAL and ELI isolates.  The vpr coding
            region of these isolates is about 18 amino acid residues longer
            than the vpr coding region of the IIIb isolates.  In HIVNL43, this
            shift is due to a single base deletion (with respect to the IIIb's)
            at position 5770.  The sequence at this position is 'atttc' in
            HIVNL43 and 'attttc' in HIVHXB2.

            The original BRU clone, sequenced by Wain-Hobson, et al. (Cell 40,
            9-17 (1985)), and the BRU portion of the pNL4-3 recombinant clone
            are different clones from the same BRU isolate.

            Two of the revisions reported in the FEATURES produced changes in
            amino acid sequences.  The revision at position 2421 changes one
            amino acid residue from 'R' to 'G' in the pol coding region.  The
            revision at positions 8995-9000 changes three amino acid residues
            from 'AHT' to 'VTP' in the nef coding region.
FEATURES             Location/Qualifiers
     LTR             1..634
                     /note="5' LTR"
     repeat_region   454..550
                     /note="R repeat 5' copy"
     prim_transcript 455..9626
                     /note="tat, rev, nef subgenomic mRNA"
     intron          744..5776
                     /note="tat, rev, nef mRNA intron 1"
     misc_feature    5743..5748
                     /note="EcoRI site of recombination"
     misc_recomb     5743..5744
                     /note="HIV-1 isolate NY5 DNA end/HIV-1 isolate LAV DNA
                     start"
     intron          6045..8368
                     /note="tat cds intron 2"
     intron          6045..8368
                     /note="rev cds intron 2"
     intron          6045..8368
                     /note="tat, rev, nef mRNA intron 2"
     LTR             9076..9709
                     /note="3' LTR"
     repeat_region   9529..9626
                     /note="R repeat 3' copy"
     polyA_signal    9602..9607
                     /note="mRNA polyadenlyation signal"
     CDS             join(5830..6044,8369..8414)
                     /note="tat protein"
                     /codon_start=1
                     /translation="MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFMTKALG
                     ISYGRKKRRQRRRAHQNSQTHQASLSKQPTSQSRGDPTGPKE"
     CDS             join(5969..6044,8369..8643)
                     /note="rev protein"
                     /codon_start=1
```

```
                /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARRNRRRR
                WRERQRQIHSISERILSTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQI
                LVESPTVLESGTKE"
CDS             790..2292
                /note="gag polyprotein"
                /codon_start=1
                /translation="MGARASVLSGGELDKWEKIRLRPGGKKQYKLKHIVWASRELERF
                AVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTIAVLYCVHQRIDVKDTKEALD
                KIEEEQNKSKKKAQQAAADTGNNSQVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVVE
                EKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPV
                HAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRM
                YSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTIL
                KALGPGATLEEMMTACQGVGGPGHKARVLAEAMSQVTNPATIMIQKGNFRNQRKTVKC
                FNCGKEGHIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQ
                SRPEPTAPPEESFRFGEETTTPSQKQEPIDKELYPLASLRSLFGSDPSSQ"
CDS             2085..5096
                /partial
                /note="pol polyprotein;  (NH2-terminus uncertain)"
                /codon_start=1
                /translation="FFREDLAFPQGKAREFSSEQTRANSPTRRELQVWGRDNNSLSEA
                GADRQGTVSFSFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMNLPGRWKPKM
                IGGIGGFIKVGQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPI
                ETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAI
                KKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKQKKSVTVLDVGDAYFSVPLD
                KDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQCSMTKILEPFRKQNPDIV
                IYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPD
                KWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYAGIKVRQLCKLLRGTKALTEVVPL
                TEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGK
                YARMKGAHTNDVKQLTEAVQKIATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWI
                PEWEFVNTPPLVKLWYQLEKEPIIGAETFYVDGAANRETKLGKAGYVTDRGRQKVVPL
                TDTTNQKTELQAIHLALQDSGLEVNIVTDSQYALGIIQAQPDKSESELVSQIIEQLIK
                KEKVYLAWVPAHKGIGGNEQVDGLVSAGIRKVLFLDGIDKAQEEHEKYHSNWRAMASD
                FNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGY
                IEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSNFTSTTVKAACWWAGIKQEFGI
                PYNPQSQGVIESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGER
                IVDIIATDIQTKELQKQITKIQNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQDNSDI
                KVVPRRKAKIIRDYGKQMAGDDCVASRQDED"
CDS             5041..5619
                /note="vif protein"
                /codon_start=1
                /translation="MENRWQVMIVWQVDRMRINTWKRLVKHHMYISRKAKDWFYRHHY
                ESTNPKISSEVHIPLGDAKLVITTYWGLHTGERDWHLGQGVSIEWRKKRYSTQVDPDL
                ADQLIHLHYFDCFSESAIRNTILGRIVSPRCEYQAGHNKVGSLQYLALAALIKPKQIK
                PPLPSVRKLTEDRWNKPQKTKGHRGSHTMNGH"
CDS             5559..5849
                /note="vpr protein"
                /codon_start=1
                /translation="MEQAPEDQGPQREPYNEWTLELLEELKSEAVRHFPRIWLHNLGQ
                HIYETYGDTWAGVEAIIRILQQLLFIHFRIGCRHSRIGVTRQRRARNGASRS"
CDS             6061..6306
                /note="vpu protein"
                /codon_start=1
                /translation="MQPIIVAIVALVVAIIIAIVVWSIVIIEYRKILRQRKIDRLIDR
                LIERAEDSGNESEGEVSALVEMGVEMGHHAPWDIDDL"
CDS             6221..8785
                /note="envelope polyprotein"
                /codon_start=1
                /translation="MRVKEKYQHLWRWGWKWGTMLLGILMICSATEKLWVTVYYGVPV
                WKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVE
                QMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
                ISTSIRDKVQKEYAFFYKLDIVPIDNTSYRLISCNTSVITQACPKVSFEPIPIHYCAP
                AGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEDVVIRSANFT
                DNAKTIIVQLNTSVEINCTRPNNNTRKSIRIQRGPGRAFVTIGKIGNMRQAHCNISRA
                KWNATLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNST
                WFNSTWSTEGSNNTEGSDTITLPCRIKQFINMWQEVGKAMYAPPISGQIRCSSNITGL
                LLTRDGGNNNNGSEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREK
```

```
                     RAVGIGALFLGFLGAAGSTMGCTSMTLTVQARQLLSDIVQQQNNLLRAIEAQQHLLQL
                     TVWGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWNNM
                     TWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKL
                     FIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPIPRGPDRPEGIEEEGGERDR
                     DRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLL
                     QYWSQELKNSAVNLLNATAIAVAEGTDRVIEVLQAAYRAIRHIPRRIRQGLERILL"
     CDS             8787..9407
                     /note="nef protein"
                     /codon_start=1
                     /translation="MGGKWSKSSVIGWPAVRERMRRAEPAADGVGAVSRDLEKHGAIT
                     SSNTAANNAACAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIH
                     SQRRQDILDLWIYHTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVEEANKGE
                     NTSLLHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFKNC"
 BASE COUNT     3421 a   1756 c   2366 g   2166 t
 ORIGIN      5' terminus of NY5 LTR

Initial Score    =    645  Optimized Score  =    645  Significance = 50.45
Residue Identity =    93%  Matches          =    645  Mismatches   =    44
Gaps             =      0  Conservative Substitutions               =     0

          10        20        30        40        50        60        70
    GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
          ||||||||||||||   ||||||  |||||||||| ||||||||||||||||||||||||||||||
          TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAA
          X       10        20        30        40        50        60

        80        90       100       110       120       130       140
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
    |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
    GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
       70        80        90       100       110       120       130

       150       160       170       180       190       200       210
    TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
    | |||| |||||||||||||| |||||  | ||||||||||||||  |||||||||| | ||||||||||||||
    TTCAAGTTAGTACCAGTTGAACCAGAGCAAGTAGAAGAGGCCAAATAAGGAGAGAAGAACAGCTTGTTACAC
    140       150       160       170       180       190       200

     220       230       240       250       260       270       280
    CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
    ||| |||||| |||||| ||||| ||||| ||| |||||||| |||| ||||| |||||||||||| |||||||
    CCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGCA
  210       220       230       240       250       260       270       280

  290       300       310       320       330       340       350       360
    TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
    |||| |||| |||||||||||||||||||||||||| ||| ||||||||||||||||| ||||||||||
    TTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGA
          290       300       310       320       330       340       350

          370       380       390       400       410       420       430
    CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
    |||||||||||| |||||||||||||| |||||||||||||| |||||||||||||||||||||||||||| |
    CTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTAC
        360       370       380       390       400       410       420

         440       450       460       470       480       490       500
    ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
    |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
    ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
      430       440       450       460       470       480       490

       510       520       530       540       550       560       570
    CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
    |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
    CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTGTGCCCGTCT
```

```
   500       510       520       530       540       550       560

    580       590       600       610       620       630       640
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
 570       580       590       600       610       620       630       640

   650       660       670       680       690     X
   CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
   ||||||||||||||||||||||||  || |||||||||  ||||||||
   CCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
         650       660       670       680       690       700       710

   CGCACGGCAAGAGGCGAGGGGCGGCG
       720       730
```

8. RAILEY-000-716.SEQ (1-696)
   AIHTLV31   Human t-cell leukemia virus type iii provirus, 5'

```
ID   AIHTLV31   standard; RNA; VRL; 660 BP.
XX
AC   K02008;
XX
DT   13-JUN-1985 (Rel. 06, Created)
DT   11-AUG-1990 (Rel. 25, Last updated, Version 1)
XX
DE   Human t-cell leukemia virus type iii provirus, 5' ltr from hxb2
XX
KW   acquired immune deficiency syndrome; long terminal repeat;
KW   provirus.
XX
OS   Human immunodeficiency virus type 1
OC   Viridae; ss-RNA enveloped viruses; Positive strand RNA viruses;
OC   Retroviridae; Lentivirinae.
XX
RN   [1]
RP   1-660
RA   Starcich B., Ratner L., Josephs S.F., Okamato T., Gallo R.C.,
RA   Wong-staal F.;
RT   "Characterization of long terminal repeat sequences of HTLV-III";
RL   Science 227:538-540(1985).
XX
CC   Acquired immune deficiency syndrome (aids) is caused by a
CC   retrovirus known by four different names, probably representing
CC   four different strains: human t-cell leukemia virus-iii (htlv-iii),
CC   aids-associated retrovirus type 2 (arv-2), aids virus, and
CC   lymphadenopathy-associated virus (lav). it is still unclear with
CC   which type of virus it is most closely associated.
CC
CC   the ltr has u3, r, and u5 regions of 453, 98, and 83 bp,
CC   respectively. this sequence has some regions homologous to human
CC   t-cell growth factor (tcgf), and the u3 region shows 83% homology
CC   with intron 1 of human gamma-interferon (gamma-if) [1]; they
CC   conclude that the regions in the htlv-iii ltr which correspond to
CC   regions in tcgf and gamma-if could be important in host cell
CC   tropism of transcriptional regulation of this virus.
XX
FH   Key             Location/Qualifiers
FH
XX
SQ   Sequence 660 BP; 160 A; 159 C; 187 G; 154 T; 0 other;

Initial Score    =    644  Optimized Score  =    645  Significance = 50.37
Residue Identity =    97%  Matches         =    645  Mismatches   =    15
```

```
Gaps            =      0  Conservative Substitutions            =      0

   X       10        20        30        40        50        60        70
   GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
     |     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TAGTAGTTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAA
   X       10        20        30        40        50        60        70

         80        90       100       110       120       130       140
   GGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
   ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGC
         80        90       100       110       120       130       140

      150       160       170       180       190       200       210
   TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TACAAGCTAGTACCAGTTGAGCCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACAC
      150       160       170       180       190       200       210

    220       230       240       250       260       270       280
   CCTGTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
   |||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||||
   CCTGTGAGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCA
    220       230       240       250       260       270       280

  290       300       310       320       330       340       350       360
   TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGA
   |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
   TTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGATATCGAGCTTGCTACAAGGGA
  290       300       310       320       330       340       350       360

         370       380       390       400       410       420       430
   CTTTCCGCTGGGCACTTTCCAGGGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGC
   |||||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||| ||||
   CTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGC
         370       380       390       400       410       420       430

       440       450       460       470       480       490       500
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGG
   ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
   ATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG
       440       450       460       470       480       490       500

     510       520       530       540       550       560       570
   CTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
   ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CTAGCTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCT
     510       520       530       540       550       560       570

    580       590       600       610       620       630       640
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   GTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGC
    580       590       600       610       620       630       640

  650         X       670       680       690
   CCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGA
   ||||||||||||
   CCGAACAGGGAC
  650       660


9. RAILEY-000-716.SE0 (1-696)
   REHIVXB2     Human T-lymphotropic virus type III (HTLV-III) 3'0
```

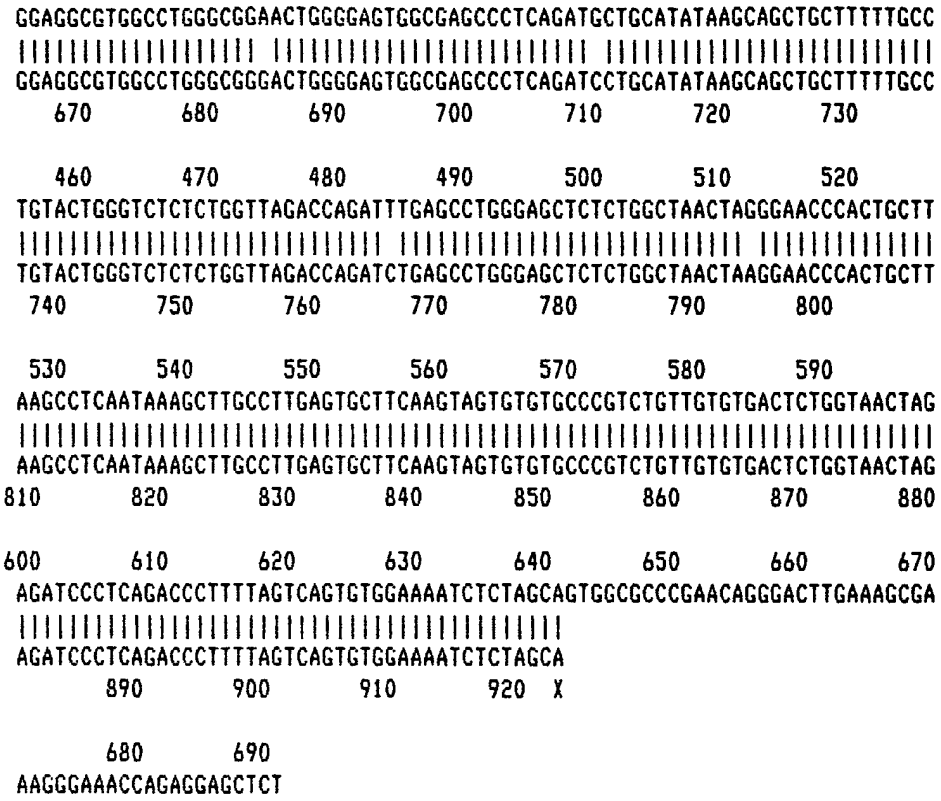
```
LOCUS       REHIVXB2      923 bp    RNA             VRL       01-JUN-1992
DEFINITION  Human T-lymphotropic virus type III (HTLV-III) 3'ORF HXB2 RNA
ACCESSION   X03187
KEYWORDS    acquired immune deficiency syndrome; long terminal repeat;
            provirus; unidentified reading frame.
SOURCE      Aids-associated retrovirus
  ORGANISM  Aids-associated retrovirus
            Viridae; ss-RNA enveloped viruses; Positive strand RNA viruses;
            Retroviridae.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Ratner,L., Starcich,B., Josephs,S.F., Hahn,B.H., Reddy,E.P.,
            Livak,K.J., Petteway,S.R.Jr., Pearson,M.L., Haseltine,W.A.,
            Arya,S.K. and Wong-staal,F.
  TITLE     Polymorphism of the 3' open reading frame of the virus associated
            with the acquired immune deficiency syndrome, human T-lymphotropic
            virus type III
  JOURNAL   Nucleic Acids Res. 13, 8219-8229 (1985)
  STANDARD  full automatic
COMMENT     *source: clone_library=lambda gtwes-lambda b; *source: clone=HXB2;

            Clone HXB2 with a termination codon at amino acid residue 124 gives
            rise to viral particles and cytopathic effects, and thus appears to
            be a fully functional clone. The N terminal portion of the 3' ORF
            protein product may include the functional region of the molecule.
            HXB2 represents an integrated proviral clone; author numbering
            refers to viral cap site at pos. +1. see x03287 - x03292.
FEATURES             Location/Qualifiers
     misc_feature    288..923
                     /note="3' LTR"
     misc_feature    288..742
                     /note="U3 sequence"
     misc_feature    743..840
                     /note="R sequence"
     misc_feature    841..923
                     /note="U5 sequence"
     CDS             1..369
                     /note="3' ORF;  (aa 1-123)"
                     /codon_start=1
                     /translation="MGGKWSKSSVIGWPTVRERMRRAEPAADGVGAASRDLEKHGAIT
                     SSNTAATNAACAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIH
                     SQRRQDILDLWIYHTQGYFPD"
BASE COUNT      249 a    207 c    262 g    205 t
ORIGIN

Initial Score    =    631  Optimized Score  =    631  Significance = 49.31
Residue Identity =    98%  Matches          =    631  Mismatches   =    10
Gaps             =      0  Conservative Substitutions               =     0

                                                    X       10        20
                                                    GGGGGACTGGAAGGGCTAATTC
                                                    |||||||||||||||||||||
   CAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTC
        240       250       260       270       280  X    290       300

         30        40        50        60        70        80        90
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
   |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
   ACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGA
      310       320       330       340       350       360       370

        100       110       120       130       140       150       160
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
    380       390       400       410       420       430       440
```

```
  170       180       190       200       210       220       230
 CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
 |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
 CAGATAAGATAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGGATGG
450       460       470       480       490       500       510       520

240       250       260       270       280       290       300       310
 ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
 ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
       530       540       550       560       570       580       590

       320       330       340       350       360       370       380
 TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
 TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAG
     600       610       620       630       640       650       660

     390       400       410       420       430       440       450
 GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
 |||||||||||||||||| |||||||||||||||||||||| |||||||||||||||||||||||||||||
 GGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGCATATAAGCAGCTGCTTTTTGCC
   670       680       690       700       710       720       730

   460       470       480       490       500       510       520
 TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
 |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||
 TGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAAGGAACCCACTGCTT
  740       750       760       770       780       790       800

  530       540       550       560       570       580       590
 AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAG
810       820       830       840       850       860       870       880

600       610       620       630       640       650       660       670
 AGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGA
 ||||||||||||||||||||||||||||||||||||||||||
 AGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA
       890       900       910       920  X

       680       690
 AAGGGAAACCAGAGGAGCTCT
```

10. RAILEY-000-716.SEQ (1-696)
    REHIVXB3 Human T-lymphotropic virus type III (HTLV III) 3'

```
LOCUS       REHIVXB3      923 bp    RNA             VRL       01-JUN-1992
DEFINITION  Human T-lymphotropic virus type III (HTLV III) 3' ORF HXB3 RNA
ACCESSION   X03188
KEYWORDS    acquired immune deficiency syndrome; long terminal repeat;
            provirus; unidentified reading frame.
SOURCE      Aids-associated retrovirus
  ORGANISM  Aids-associated retrovirus
            Viridae; ss-RNA enveloped viruses; Positive strand RNA viruses;
            Retroviridae.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Ratner,L., Starcich,B., Josephs,S.F., Hahn,B.H., Reddy,E.P.,
            Livak,K.J., Petteway,S.R.Jr., Pearson,M.L., Haseltine,W.A.,
            Arya,S.K. and Wong-staal,F.
  TITLE     Polymorphism of the 3' open reading frame of the virus associated
            with the acquired immune deficiency syndrome, human T-lymphotropic
            virus type III
  JOURNAL   Nucleic Acids Res. 13, 8219-8229 (1985)
```

```
  STANDARD  full automatic
COMMENT     *source: clone_library=lambda gt wes-lambda b; *source: clone=HXB3;

            HXB3 represents an integrated proviral clone; see x03187 - x03190;

            author numbering refers to viral cap site at pos. +1.
FEATURES             Location/Qualifiers
     misc_feature    288..923
                     /note="3' LTR"
     misc_feature    288..742
                     /note="U3 sequence"
     misc_feature    743..840
                     /note="R sequence"
     misc_feature    841..923
                     /note="U5 sequence"
     CDS             1..618
                     /note="3' ORF;  (aa 1-206)"
                     /codon_start=1
                     /translation="MGGKWSKSSVVGWPAVRERMRRAEPAADGVGAASRDLEKHGAIT
                     SSNTAANNAACAWLEAQEEEKVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIH
                     SQRRQDILDLWIYHTQGYFPDWQNYTPGPGIRYPLTFGWRYKLVPVEPEKLEEANKGE
                     NTSLLHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFKNC"
BASE COUNT      252 a    208 c    260 g    203 t
ORIGIN

Initial Score    =    626  Optimized Score  =    626  Significance = 48.90
Residue Identity =    97%  Matches          =    626  Mismatches   =    15
Gaps             =      0  Conservative Substitutions               =     0

                                                    X       10        20
                                                    GGGGGACTGGAAGGGCTAATTC
                                                    ||||||||||||||||||||||
   CAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTC
        240       250       260       270       280  X    290       300

         30        40        50        60        70        80        90
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ACTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGA
      310       320       330       340       350       360       370

      100       110       120       130       140       150       160
   ACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGC
   |||||||||||||| |||||| | |||||||||||||||||||||||| |||||||||||||||||||||||
   ACTACACACCAGGACCAGGGATAAGATATCCACTGACCTTTGGATGGCGCTACAAGCTAGTACCAGTTGAGC
    380       390       400       410       420       430       440

     170       180       190       200       210       220       230
   CAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
   |||| ||| |||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
   CAGAGAAGTTAGAAGAAGCCAACAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGAATGG
  450       460       470       480       490       500       510       520

  240       250       260       270       280       290       300       310
   ATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
   ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   ATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGC
         530       540       550       560       570       580       590

         320       330       340       350       360       370       380
   TGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGCACTTTCCAG
   |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||||
   TGCATCCGGAGTACTTCAAGAACTGCTGATATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCAG
        600       610       620       630       640       650       660

        390       400       410       420       430       440       450
```


GGAGGCGTGGCCTGGGCGGAACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCC
GGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGCATATAAGCAGCTGCTTTTTGCC
670 680 690 700 710 720 730
460 470 480 490 500 510 520
TGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
TGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAAGGAACCCACTGCTT
740 750 760 770 780 790 800
530 540 550 560 570 580 590
AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAG
AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAG
810 820 830 840 850 860 870 880
600 610 620 630 640 650 660 670
AGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGA
AGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA
890 900 910 920 X
680 690
AAGGGAAACCAGAGGAGCTCT